TOWARDS THE EXTENSION OF THE KNOWLEDGEBASE TO FURTHER THE UNDERSTANDING AND MODELLING OF DRIVER BEHAVIOUR

by

Pieter Poolman

DISSERTATION (Volume 2)

Presented to the Faculty of Engineering of the University of Stellenbosch in Partial Fulfilment for the Degree of

Ph.D. (CIVIL ENGINEERING)

Promoter:
Dr. C.J. Bester

STELLENBOSCH
APRIL 2003
# TABLE OF CONTENTS

## APPENDIX A LITERATURE REVIEW OF DRIVER BEHAVIOUR

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Culture of Driving</td>
<td>1</td>
</tr>
<tr>
<td>A.1.1 Summary</td>
<td>1</td>
</tr>
<tr>
<td>A.1.2 Relationship between personality and driver behaviour</td>
<td>7</td>
</tr>
<tr>
<td>A.1.3 Culture of driving: individuals in a social environment</td>
<td>9</td>
</tr>
<tr>
<td>A.1.4 The effects of enforcement on compliance: perceived and objective risk</td>
<td>18</td>
</tr>
<tr>
<td>A.1.5 Can traffic law enforcement permanently reduce the number of accidents?</td>
<td>20</td>
</tr>
<tr>
<td>A.1.6 An evaluation of the effectiveness of police intervention on speeding</td>
<td>22</td>
</tr>
<tr>
<td>A.1.7 Causes and manifestations of aggression in car driving</td>
<td>26</td>
</tr>
<tr>
<td>A.1.8 Sensation seeking and risky driving</td>
<td>30</td>
</tr>
<tr>
<td>A.1.9 Objective and subjective risk, risk perception and driver behaviour</td>
<td>38</td>
</tr>
<tr>
<td>A.1.10 Drivers’ perception of specific hazards on the road</td>
<td>40</td>
</tr>
<tr>
<td>A.1.11 Drivers’ perception of risk in automobiles</td>
<td>43</td>
</tr>
<tr>
<td>A.1.12 Accident risk and perceived driving ability of young and older drivers</td>
<td>44</td>
</tr>
<tr>
<td>A.1.13 Driving skills: experience and expertise</td>
<td>47</td>
</tr>
<tr>
<td>A.1.14 Theory of goal weighting in driver behaviour</td>
<td>49</td>
</tr>
<tr>
<td>A.1.15 Risk homeostasis theory defended by Wilde (1988)</td>
<td>55</td>
</tr>
<tr>
<td>A.1.16 Reasons for safety measures deemed unsuccessful by risk homeostasis theory</td>
<td>61</td>
</tr>
<tr>
<td>A.1.17 Implications of the zero-risk theory of driver behaviour</td>
<td>65</td>
</tr>
<tr>
<td>A.1.18 Fuller’s threat avoidance theory of driving behaviour</td>
<td>67</td>
</tr>
<tr>
<td>A.1.19 Decision-making, risky behaviour and bounded rationality</td>
<td>71</td>
</tr>
<tr>
<td>A.1.20 Mental load, stress and risk exposure in traffic</td>
<td>75</td>
</tr>
<tr>
<td>A.1.21 Variations of the level of vigilance and of behavioural activities during automobile driving</td>
<td>78</td>
</tr>
<tr>
<td>A.1.22 Accident causation and decision-making in emergency situations</td>
<td>80</td>
</tr>
<tr>
<td>A.2 Driver Training</td>
<td>83</td>
</tr>
<tr>
<td>A.2.1 Summary</td>
<td>83</td>
</tr>
<tr>
<td>A.2.2 Modelling young and novice drivers’ behaviour</td>
<td>84</td>
</tr>
</tbody>
</table>
A.2.3 Learner drivers’ practice, socio-economic standing and lifestyle .......... 86
A.2.4 Conflicting goals of advanced (skid) training ................................ 88
A.2.5 Crash involvement rates of novice drivers of all ages ....................... 90
A.2.6 Difficulties in a driver’s process of learning ................................. 91

A.3 Drivers’ Perception of Speed .................................................. 94
  A.3.1 Summary ............................................................................... 94
  A.3.2 Effect of driving speed on reaction time ................................... 94
  A.3.3 Estimation of vehicular velocity under time limitation and restricted condition of observation ..................................................... 96
  A.3.4 Scaling of relative velocity between vehicles ............................. 97
  A.3.5 Knowledge of oncoming car speed as determiner of driver’s passing behaviour ................................................................. 99
  A.3.6 Choice of time-headway in car following and the role of time-to-collision information in braking ............................................ 100
  A.3.7 Modelling car-following behaviour ........................................ 103
  A.3.8 Close-following drivers on two-lane highways .......................... 104
  A.3.9 New systems: new behaviour? .............................................. 105

A.4 Lane keeping for Straight and Curve Driving .................................. 107
  A.4.1 Summary ............................................................................... 107
  A.4.2 Vehicle control in straight lane keeping ..................................... 107
  A.4.3 Vehicle control during curve driving ....................................... 109
  A.4.4 Choice of speed and steering behaviour during curve driving ....... 111
  A.4.5 Fixation during car driving .................................................... 112
  A.4.6 Maintaining lane position with peripheral vision during in-vehicle tasks ................................................................. 113
  A.4.7 Application of approximate time-to-line-crossing (TLC) during car driving ...................................................... 115

A.5 Drivers’ Visual Search ............................................................... 116
  A.5.1 Summary ............................................................................... 116
  A.5.2 What attracts attention when driving? Drivers report ................ 117
  A.5.3 Eye-head dynamics, visual search and risk levels ....................... 121
  A.5.4 Strategies of visual search by drivers – fixation characteristics ... 123
  A.5.5 Strategies of visual search by drivers – recent findings ............... 124
  A.5.6 Reading the road - visual attention and saliency ....................... 126
  A.5.7 Models of visual attention ...................................................... 127
  A.5.8 Visual attention problems of older drivers ............................... 128
A.5.9 Driving experience and visual attention in car following when looking at in-car targets ................................................................. 130
A.5.10 Safety implications for using mobile (cellular) telephones while driving ................................................................. 131
A.5.11 Driving experience and time-sharing during in-car tasks on roads of different width ................................................................. 132

A.6 Dimensions of aberrant driver behaviour ........................................................................................................... 134
   A.6.1 Summary ................................................................................................................................................ 134
   A.6.2 Errors and violations: dimensions of aberrant driver behaviour ................................................................. 134
   A.6.3 A model and proposed analysis technique for errors in driving tasks ................................................................. 136
   A.6.4 A basic driver error: late detection .............................................................................................................. 138
   A.6.5 Error recognition, correction and signalling .................................................................................................. 140
   A.6.6 Variable errors setting a limit to adaptation .................................................................................................. 143
   A.6.7 Prospects for technological countermeasures against driver fatigue ................................................................. 144
   A.6.8 Designing for transportation safety in the light of perception, attention and mental models ................................................................. 145
   A.6.9 Current and future perspectives on error analyses and traffic behaviour ................................................................. 147

A.7 Driver Behaviour Modelling ........................................................................................................................................ 150
   A.7.1 Summary ................................................................................................................................................ 150
   A.7.2 A Review of driver models (Ranney, 1994) .................................................................................................. 151
   A.7.3 Transport automation and driver behaviour modelling .................................................................................. 170
   A.7.4 Example of an information-processing driver model based on artificial neural networks ................................................................. 175
   A.7.5 Example of an information-processing driver model based on pre-defined rules and algorithms ................................................................. 178
   A.7.6 Autonomous driver model in driving simulation environments ........................................................................ 180

A.8 Drivers’ Emotions ................................................................................................................................................ 183
   A.8.1 Summary ................................................................................................................................................ 183
   A.8.2 Personality, driver stress and accident proneness .......................................................................................... 184
   A.8.3 Control motivation and decision making of drivers ...................................................................................... 186
   A.8.4 Threat appeals and reckless driving ............................................................................................................ 189
   A.8.5 Threefold system of driver behaviour (James, 1999) ...................................................................................... 192

APPENDIX B THE BRAIN: PHYSIOLOGICAL, ANATOMICAL, DEVELOPMENTAL, AND NEUROPSYCHOLOGICAL ISSUES ................. 198
B.1 Unifying Brain Dynamics and Modularity ............................................................................. 198
  B.1.1 Overview .................................................................................................................. 198
  B.1.2 Complementary expectation learning and matching during ‘what’ and ‘where’ processing ................................................................................................................. 200
  B.1.3 More on motor control .............................................................................................. 203
  B.1.4 Complementary attentive-learning and orienting-search ........................................ 209
  B.1.5 Complementary additive and subtractive intrastream processing .............. 212
  B.1.6 Factorisation of pattern and energy: ratio processing and synchrony .............. 213
  B.1.7 Cognitive-emotional interactions and attentional blocking ......................... 215
  B.1.8 Rate-invariant speech and language understanding ............................................. 217
  B.1.9 Summary .................................................................................................................. 219

B.2 Current Opinion on Selected Topics in Cognitive Neuroscience ................................. 220
  B.2.1 Overview .................................................................................................................. 220
  B.2.2 Aspects of the functioning of neurotransmitters and neurons .......................... 222
  B.2.3 Aspects of brain development and structure ......................................................... 232
  B.2.4 Learning and representation in speech and language ......................................... 238
  B.2.5 Plasticity and organisation of action ................................................................. 244

B.3 Social Cognition .............................................................................................................. 260
  B.3.1 Overview .................................................................................................................. 260
  B.3.2 Social cognition, modularity and innateness ......................................................... 261
  B.3.3 Social judgment of faces in the amygdala .............................................................. 263
  B.3.4 Social reasoning and decision making in the ventromedial prefrontal cortex .... 265
  B.3.5 Representing the minds of others ................................................................. 266
  B.3.6 Empathy and simulation in the somatosensory cortices ..................................... 267
  B.3.7 Summary .................................................................................................................. 268

B.4 The Underpinnings of Emotions .................................................................................... 270
  B.4.1 Introduction .............................................................................................................. 270
  B.4.2 Cognition, emotion and the computer metaphor ................................................. 272
  B.4.3 Appraising emotions ............................................................................................. 279
  B.4.4 Nature of emotions ............................................................................................... 282
  B.4.5 Anatomy and physiology of the fear system ....................................................... 285
  B.4.6 Emotions and memory systems of the brain ........................................................ 290
  B.4.7 Stress and loss of memory ..................................................................................... 294
B.4.8 Model of emotions

B.5 Explaining Consciousness and Attention

B.5.1 Theory of consciousness (Dennett, 2001)

B.5.2 A Neural Global Workspace Model for Conscious Attention (Newman et al., 1997)

B.5.3 Temporal binding and neural correlates of sensory awareness (Engel and Singer, 2001)

APPENDIX C CONNECTIONIST MODELLING OF COGNITIVE PROCESSES

C.1 Overview of Connectionism

C.1.1 Challenges of connectionist modelling

C.1.2 Capabilities of connectionist modelling

C.2 Principles of Connectionism

C.2.1 Basics of information processing

C.2.2 Attraction of parallel distributed processing for modelling cognition

C.2.3 Selective history of early connectionism

C.2.4 Architecture and properties of a pattern associator

C.2.5 Architecture and properties of an autoassociator

C.2.6 Training a multi-layer network with an error signal – hidden units and back-propagation

C.2.7 Competitive networks

C.2.8 Recurrent networks

C.3 Applications of Connectionist Models

C.3.1 Reading aloud

C.3.2 Language acquisition

C.3.3 Cognitive development

C.3.4 Lesioning connectionist networks

C.3.5 Mental representation: rules and symbols in connectionist networks

C.3.6 Network models of brain function

C.3.7 Evolutionary connectionism
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1</td>
<td>Range ($R$) versus range-rate ($R_{dot}$) phase space (Fancher et al., 2001)</td>
<td>103</td>
</tr>
<tr>
<td>A.2</td>
<td>Classification of selected driving tasks</td>
<td>164</td>
</tr>
<tr>
<td>B.1</td>
<td>The LAMINART model of bottom-up, top-down and horizontal interactions in the lateral geniculate nucleus (LGN) and visual areas V1 and V2. (Numbers on the right refer to cortical layers) (Grossberg, 2000)</td>
<td>202</td>
</tr>
<tr>
<td>B.2</td>
<td>The VITE circuit model – thick connections represent the kinematic feedback control aspect of the model, whereas thin connections represent additional compensatory circuitry. (CBM: assumed cerebello-cortical input to IFV stage; c.s.: central sulcus; GO: scaleable gating signal; IFV: inertial force vector; i.p.s.: intraparietal sulcus; OFPV: outflow force and position vector; OPV: outflow position vector; SFV: static force vector; $\alpha$: alpha motor neuron; $\gamma^d$: dynamic gamma motor neuron; $\gamma^s$: static gamma motor neuron; la: type la afferent fibre; II: type II afferent fibre; +: excitation; -: inhibition; $\times$: multiplicative gating; $\hat{f}$: integration) (Grossberg, 2000)</td>
<td>204</td>
</tr>
<tr>
<td>B.3</td>
<td>Schematic of the multiple-paired forward-inverse model with $n$ paired modules. (dotted lines: training signals; $\times$: signal multiplication) (Wolpert et al., 1998)</td>
<td>207</td>
</tr>
<tr>
<td>B.4</td>
<td>Example of a search for a recognition code within an adaptive resonance theory (ART) learning circuit. (The process is described in detail by Carpenter and Grossberg (1991))</td>
<td>210</td>
</tr>
<tr>
<td>B.6</td>
<td>The diagram on the left shows a slice through one hemisphere of the brain, roughly between the ears and through the somatosensory cortex. On the map of the body surface some areas, like face and fingers, have a much larger cortical representation than others, distorting the map. The diagram on the right shows the map of the muscles in the motor cortex on a slice of the other hemisphere and at a more anterior position. (Gregory, 1987)</td>
<td>245</td>
</tr>
<tr>
<td>B.7</td>
<td>Summary of connection between visual cortical areas (V1, V2) and primary motor cortex (M1). (Cb = cerebellum; Cing = cingulated cortex; PN = pontine nuclei; pSMA = pre-SMA; PM = premotor cortex; VPLo = oral portion of ventral posterior lateral thalamus) (Schwartz, 1994)</td>
<td>251</td>
</tr>
<tr>
<td>B.8</td>
<td>A contemporary map of some cortical functions (LeDoux, 1998)</td>
<td>276</td>
</tr>
<tr>
<td>B.9</td>
<td>The branching tree of brain evolution (LeDoux, 1998)</td>
<td>283</td>
</tr>
</tbody>
</table>
Figure B.10 Organisation of information-processing pathways in the amygdala (LeDoux, 1998) ................................................................. 286
Figure B.11 The low and high roads to the amygdala (LeDoux, 1998) .................................................................................. 287
Figure B.12 Brain pathways of defence (LeDoux, 1998) ........................................................................................................... 288
Figure B.13 The amygdala as the hub in the wheel of fear (LeDoux, 1998) ................................................................. 289
Figure B.14 Brain systems of emotional memory and memory of emotion (LeDoux, 1998) ........................................... 290
Figure B.15 Frontal cortex areas involved in aspects of working memory (LeDoux, 1998) .................................. 297
Figure B.16 The influence of the amygdala on sensory areas of cortex is greater than the influence of the same areas on the amygdala (LeDoux, 1998) ................................................................. 299
Figure B.17 Some cortical outputs of the amygdala and their function (LeDoux, 1998) .................................................. 300
Figure B.18 Some neural ingredients of a conscious emotional experience (LeDoux, 1998) ........................................... 302
Figure B.19 The wiring diagram of the model of the thalamus-nRt-cortex complex. Input $I_j$ is sent both to the thalamic relay cell $T_j$ and the inhibitory interneuron $IN_j$ which also feeds to $T_j$. Output from $T_j$ goes to the corresponding cortical cell $C_j$ which returns its output to $T_j$. Both the axons $T_jC_j$ and $C_jT_j$ send axon collaterals to the corresponding nRt cell $N_j$. There is axonal output from $N_j$ to $IN_j$, as well as collaterals to neighbouring nRt cells, and dendro-dendritic synapses between nRt cells. (Taylor and Alavi, 1993) .................................................................................. 310
Figure B.20 ‘Wagon wheel’ model of CNS systems contributing to global attention and conscious perception. (A1: primary auditory area; gc: ‘closed’ nRt gate; go: ‘open’ nRt gate; PFC: prefrontal cortex; SI: primary somatosensory area; V1: primary visual cortex) (Newman et al., 1997) .................................................................................. 312
Figure C.1 Three types of neuron: from left to right, a spinal motoneuron, a hippocampal pyramidal cell, and a Purkinje cell of the cerebellum; the last figure is a computational unit in a connectionist model (McLeod et al., 1998) .................................................................................. 333
Figure C.2 The layered structure of information processing in the brain and connectionist models (McLeod et al., 1998) .................................................................................. 334
Figure C.3 A pattern association network in the nervous system (a), and in conventional connectionist format (b) (McLeod et al., 1998) .................................................................................. 346
Figure C.4 An eight unit autoassociator (McClelland and Rumelhart, 1985) .................................................................................. 349
Figure C.5 Recall of a complete image by an autoassociator given a noisy or partial clue (Hertz et al., 1991) .................................................................................. 351
Figure C.6 A simple competitive learning network with inhibitory connections in conventional connectionist format (a), and in the nervous system without inhibitory connections (b) (McLeod et al., 1998) .................................................................................. 355
Figure C.7 A simple recurrent network (SRN) (McLeod et al., 1998) ... 358
Figure C.8 An arbitrary mapping problem with input patterns (+s and −s) and attractors (dots) (a), solved by a recurrent network creating two basins separated by the curved line (b) (McLeod et al., 1998) ... 359
Figure C.9 A typical ‘2-route’ model of reading aloud (McLeod et al., 1998) ... 361
Figure C.10 The dual-route model for the English past tense (Pinker and Prince, 1988) ... 365
Figure C.11 A simplified version of the network architecture for lexical development (McLeod et al., 1998) ... 367
Figure C.12 A simple ‘phrase structure’ for the sentence ‘The boy kicked the ball’. (NP = noun phrase; VP = verb phrase; Art = article) (McLeod et al., 1998) ... 369
Figure C.13 Network used by Mareschal et al. (1995) to simulate the development of object permanence ... 373
Figure C.14 The retinal input and prediction of the position of an object by the network of Mareschal et al. (1995) ... 373
Figure C.15 A network model which performs the Stroop task (Cohen et al., 1990) ... 385
Figure C.16 Position and detailed structure of the hippocampus (McLeod et al., 1998) ... 389
Figure C.17 A schematic representation of connections within the hippocampus (McLeod et al., 1998) ... 391
Figure C.18 A schematic representation of the flow of information between the hippocampal system and neocortex – thick lines above the cell bodies represent dendrites, and thinner lines with arrow heads the axons. (Δ: pyramidal cell bodies; ○: dentate granule cells; mf: mossy fibre; pp: perforant path; rc: recurrent collaterals of CA3 cells; PHG: parahippocampal gyrus) (McLeod et al., 1998) ... 391
Figure C.19 A neural network simulation of the hippocampus (McLeod et al., 1998) ... 393
Figure C.20 Performance of a network simulating hippocampal episodic memory retrieval – a comparison is made for the complete network (entorhinal cortex) and in two earlier stages, CA3 and CA1 (McLeod et al., 1998) ... 394
Figure C.21 The architecture of VisNet (McLeod et al., 1998) ... 397
APPENDIX A

LITERATURE REVIEW OF DRIVER BEHAVIOUR

A.1 Culture of Driving

A.1.1 Summary

Personality researchers and theorists are approaching consensus on the basic structure and constructs of personality. Much evidence has been accumulated relating different personality dimensions to deficits in driving performance. Studies show that a behaviour pattern of competitive achievement-striving, an exaggerated sense of time urgency, aggressiveness, and impatience is significantly related to the incidence of traffic accidents and violations. Results also demonstrate a significant inverse relation between conscientiousness and driving accident involvement.

Understanding the interrelationships between individual behaviour and that of a population of drivers is important in many areas of transportation, traffic control, and safety management. The social sciences suggest many mechanisms of social influence, for example: learning through reward and punishment; compliance with leaders or other authoritative figures; conforming to norms, laws, and regulations; acceptance of persuasive messages. What we think other people think or might do (related to “Theory of Mind” in psychology) is important input for our own decision process. The fact that decisions are an ongoing process requires modelling the progressive changes in the value or utility of the outcomes of previous decisions as these are considered in future choices. In modelling traffic behaviour one should recognise that road users’ actual behaviour is determined by their view of the traffic system and that this view encompasses a complex structure of beliefs related to other systems. The efficacy of engineering designs, administrative procedures, or technical devices aimed at improving safety may also depend, in part, on drivers’ acceptance and willingness to utilise them as intended.

Police goals of increasing apprehension of traffic law violators and increasing compliance with these laws are generally viewed as being contradictory. This is because for increased apprehension, it is best for the police to reduce their perceived presence, and, thus reduce the
driver's perceived risk of apprehension. However, to increase compliance, it is best to increase the perceived risk of apprehension, which is typically brought about by increasing the police conspicuity and visibility. The major determinant in the compliance behaviour is not necessarily what the police do, but rather, how the road users perceive it. A number of studies show that the effects of police enforcement are transient and local in nature. Road users adapt to changes in the perceived probability of encountering enforcement. If enforcement is encountered, violation rates drop momentarily, but return to their previous level once the site of enforcement has been passed or as time passes since the last enforcement was encountered. Violations and road accidents can be permanently reduced if enforcement is not reduced once it is successful. Furthermore, general attitude towards exceeding posted speed limits, and beliefs about the consequences of such speeding, are important determinants of intentions to speed and actual speeding behaviour, as are social facilitation effects based on comparison of one's own speed with others on the road. It has been shown that speeding is still perceived as socially acceptable, with people believing that most others speed; that their peers would approve of their behaviour; and that it is unlikely to result in prosecution or an accident.

Aggressive behaviour among motorists has become a common social problem in recent years. Potentially angering driving-related situations include hostile gestures, illegal driving, police presence, slow driving, discourtesy, and traffic obstructions. In general, results from studies revealed that drivers who have an accident record score high on aggression and alcohol abuse and low on extraversion personality dimensions. These drivers also drive more quickly on clear roads, move about more, and overtake in traffic. It is also clear that an enclosed vehicle provide a sense of anonymity to its occupants, which facilitates aggression on the highway.

Since the 1970s, traffic safety researchers have observed and examined the relationship between sensation seeking and risky driving (for example, driving while impaired, speeding, following too closely), as well as its consequences (for example, violations and collisions). The relationship is stronger with observed or reported driving behaviour than with traffic violations or collision involvement. In addition, men are more often involved in sensation seeking and risky driving than women are. There is also growing evidence that sensation seeking may also moderate the manner in which drivers respond to other factors such as alcohol impairment and perceived risk. However, the exact psychological mechanisms through which sensation seeking influences driving and the nature of the biological bases of sensation seeking require further research.
It is often assumed that human behaviour is more directly determined by subjective estimates of risk than by objective risk. While this may be the case for events about which there is little experience such as nuclear accidents, it is probably not the case for everyday events such as the risks encountered on the road. Difficulties in understanding probabilistic processes, biased media coverage, misleading personal experiences, and the anxieties generated by life's gambles, frequently lead individuals to deny uncertainty, misjudge risks, and maintain unwarranted confidence in judgments of fact. In general, human information processing is hindered by biases and limitations which affect subjective evaluation of probabilities. Perceptions are often resistant to change: new evidence which is consistent with initial perceptions is accepted, while that which is contrary is dismissed.

Drivers modify their behaviour according to the risk they perceive. It would be valuable to identify those aspects of the environment which road users take into account and how much each aspect contributes to the perceived risk. Drivers appeared to view the road as a fixed landscape, which it is their job to negotiate successfully. However, drivers cannot talk generally about road hazards, and appear to rely substantially on visual imagery, suggesting they only think of hazards as specific events in the past. When road hazards are mentioned, they usually have to do with the behaviour of other road users. One possible reason for the low level of responses concerning specific ‘road environment’ hazards is inadequate learning. Furthermore, although improvements of safety devices in automobiles have reduced the effects of road accidents substantially, many people seem to think that these improvements will automatically produce safety without effort by the driver or passengers. It is clear that the level of knowledge about some automotive hazards and the understanding of relevant physical principles are relatively poor.

Young drivers are significantly over represented among all drivers involved in traffic accidents and fatalities. Researchers found that inexperienced drivers are more likely to be involved in accidents due to improper directional and speed control of their vehicles and for taking longer to respond to potentially dangerous traffic circumstances. Another interpretation is that young drivers take more risks than older drivers do. To gain experience, young drivers must spend time behind the wheel, but in doing so they increase their chances of being involved in an accident because of their inferior skills. As to basic driving skills, it is possible to lengthen driver education and make the licence examination more stringent and, if necessary, to delay licensing until basic car control skills have been acquired. However, neither skills nor knowledge can save one from risky driving and accidents.
After passing a driving test and obtaining a licence, the ordinary motorist is left largely to his/her own devices during the subsequent years of experience on the road. The question is what happens to driving skills during this extended period of unsupervised experience. Both theoretically and practically, it is important to know why experience improves some driving skills, but allowing bad habits to develop in others. Studies suggest that the nature of feedback in unsupervised driving may be crucial. Where feedback is good, simple experience may bring expertise, but where feedback is poor, skills may fail to improve or even deteriorate once explicit tuition is removed. For example, experienced drivers are relatively poor in aspects of mirror checking and setting a safety margin, for all of which immediate feedback from the driving situation is usually weak. These then may be the sorts of skill most dependent on the supplementary feedback that can be provided by a tutor; and therefore the skills most likely to deteriorate once a tutor is no longer present. By contrast, experienced drivers accelerate and change gears relatively quickly, and have efficient, smooth, and consistent steering paths.

It can be argued that human behaviour is intrinsically goal-directed. An interpretation concerning which goals seem important to pursue in the everyday traffic environment is an alternative explanation of why some aspects of driving improve with experience, while others equally clearly do not. A theory of goal choice relate topics such as driver skill development, risk-taking and individual differences by dealing with the flexibility of real life, the role of practice, and the question of distractibility and the variable efficiency in a process of weighting alternative goals in their competition for control of behaviour. Even well-practised behaviour is still goal-directed in that familiar stimuli are linked not just to associated actions, but also to associated next goals. Each current state provides such strong activation to the familiar next-state as almost to ensure that it will gain control of behaviour.

Questions as to whether, in which manner and to what extent road users alter their behaviour and their accident risk in response to legal, educational and technological interventions and innovations attract a considerable amount of attention. Risk homeostasis theory (RHT) has been formulated to answer these questions. This has met with both criticism and support. From RHT it follows that the population traffic-accident rate will not drop as a result of regulations mandating the use of seat belts or motorcycle helmets, limits to speed or blood-alcohol levels, vehicle manufacturing standards, the training of road-user skills, or indeed following any other intervention which does not reduce the target level of traffic accident risk. These interventions, effective in changing the frequency of specific behaviours in traffic to which they are addressed, lead road users to display other (for example, non-regulated) risky behaviours in road traffic at a higher frequency. However,
accident countermeasures which take the form of reducing the target level of risk have already demonstrated their effectiveness in drastically reducing the accidents rate per person in both road traffic and industry.

Risk homeostasis theorists cite as evidence in favour of the theory the failure of specific safety measures to reduce accidents, however, this assertion is legitimate only if the mediating behavioural changes can be identified. The reason for this is that safety measures may fail for a wide range of reasons quite unrelated to risk homeostasis theory. In general, if a safety measure is not implemented, or is implemented incorrectly, then it is unlikely to operate effectively. One must distinguish between the ineffectiveness of a safety measure due to a behavioural change following correct implementation and a failure of implementation itself. Furthermore, people do not have the required sensitivity to low probability events such as accidents. In addition, it has been found that context effects are important for risk perception. Thus, risk perception is not simply related to the frequency of accident casualties as risk homeostasis theorists assert. Any theoretical position which relies only on the frequency of accident casualties as its crucial variable, will remain unsatisfactory because people’s perceptions are determined by a variety of factors, including perceived controllability and capacity for catastrophic potential.

Automation of the driving task and avoidance learning make it possible that most of driving eventually becomes a habitual activity based largely on automatised control of safety margins in partial tasks. No consideration is normally given to risks, as explained by the zero-risk theory. However, in most situations, drivers know what they should not do if they want to avoid a certain, or an almost certain, accident. And, whenever necessary, they can focus their conscious attention on estimates as to whether, say, a gap is sufficiently wide to permit merging, or a following distance is sufficient, or the outer lane is free for safe overtaking. Similarly, they can switch to conscious, controlled information processing at any other level of the driving task. Furthermore, drivers are quick to use any opportunity provided by new outlets for their motives if the system changes (environmental, vehicular or statutory). The key to effective safety countermeasures is thus to prevent drivers from changing their behaviour in response to system modifications: that is, to prevent them from satisfying their motives. The psychological implication of the zero-risk theory is that one needs behavioural restrictions in the traffic system: quite contrary to what the proponents of the risk homeostasis theory state.

The threat-avoidance model presupposes that drivers opt for zero risk of accident (i.e. it assumes that such events are always aversive) but that they may make anticipatory avoidance, competing or
delayed avoidance responses depending on a variety of factors. These include the rewards and punishments associated with each kind of response, the accuracy with which discriminative stimuli are recognised, the subjective probability of a potential aversive stimulus, the effectiveness of avoidance responses when made, and the driver’s arousal condition. The driver’s previous experience is paramount in determining the nature and values of each of these variables. The extinction of learned associations between discriminative stimuli and potential aversive stimuli and in addition, extinction of anticipatory avoidance responses themselves through lack of reinforcement, may account for shifts away from safety in driver behaviour. These shifts are observable until punishing consequences ensue to restore both associations and contingencies.

Risk-taking, and ultimately accidents, in traffic may be defined as the result of discrepancy between the road user’s self-imposed demands of speed and accuracy and the actual availability of human aptitudes and psychological resources to meet those demands. Decisions in traffic often have to be made on the basis of predicted events, rather than in real-time. Since road traffic has a considerable inertia, predictions are useful. These predictions may be less or more accurate. One of the main topics of the theory of decision-making under uncertainty is ‘bounded rationality’, which is a psychological alternative for the assumptions concerning the normative concepts of how decisions should be taken. In making a decision, the driver appears to give a high value to information which confirms hypotheses derived from initial data, and tends to neglect information which disconfirms the hypothesis. Thus, a driver’s decision-making takes into account a very narrow range of possibilities and is biased by initial impressions.

‘Stress’ behind the wheel has become a concern for both behavioural scientists who study traffic, and the public. Stress, in the sense of mental overload, is seen not only as a subjective and limiting condition about which many drivers complain, but also as a possible cause of disturbances in the traffic system and even of many traffic accidents. Strain is regarded as the process of putting one’s capacities to use in the course of ‘normal’ work activities. Models of stress and strain are based on the assumption that there is a lack of equilibrium between individuals and their environment, which results from a discrepancy between demands and possible ways of dealing with situations. An optimal amount of strain contributes to optimal task execution. If there is a lack of balance between the activation level and the performance capacity of a driver, or if the demands which traffic makes on performance produce tensions, then there is the risk of an accident, in which the necessity to make decisions under time pressure and using limited mental capacity plays an important role. Strain can also lead to inadequate emotional reactions such as fear or anger, which require internal regulation.
There exist two types of behavioural activities: those necessary for the performance of the task and non-specific activities which are not directly imposed by the task. The monotony of the working situation induces the increase of behaviours which are irrelevant to the primary task. These behavioural variations are interpreted as being an indication of an individual's level of arousal. The occurrence of these activities is a defence mechanism which enables the maintenance of performance or of a stable level of output. However, they are not always efficient. In addition, an individual's self-awareness of which behavioural activities are precursory signs for him/her of an increase in low vigilance, does not allow one to assume that he/she will be able to notice them while engaged in a monotonous task. Similarly, knowing which activities can be reactivating does not allow one to know if the driver will be in a position to voluntarily make use of them to reactivate himself/herself, if necessary.

Accidents are preceded by long histories containing multitudes of events. These antecedent events can be classified into at least four groups which occur in this order: failure types; psychological precursors; unsafe acts; and breakdown of defences. In the event of an emergency, studies show that drivers rarely attempt sideways avoidance manoeuvres, but that their usual course of action is to brake, with the result that the wheels lock leaving other manoeuvres impossible. In general, drivers' knowledge of actual avoidance possibilities is vague and irregular. Given the effects of stress and temporal constraints, one might assume that drivers do not rely on refined perceptual judgments for taking collision avoidance decisions, but tend to follow simplified strategies, resorting to braking as their initial reaction.

A.1.2 Relationship between personality and driver behaviour

Type A behaviour pattern

Friedman and Rosenman (1959) described the Type A behaviour pattern. Much evidence has since been accumulated which shows that Type A individuals are characterised by competitive achievement-striving, an exaggerated sense of time urgency, aggressiveness, and impatience (Jenkins, 1976). The relative absence of Type A characteristics constitute the Type B behaviour pattern. These latter individuals exhibit an overt behaviour pattern that is more relaxed and easygoing.
Abrahams and Birren (1973) demonstrated that the reaction time performance deficit found in persons with coronary heart disease was also present in pathology-free Type A individuals. In fact, Type A subjects had greater response latencies than Type B subjects in both simple and complex choice-reaction time performance tasks. An interpretation of these results, as suggested by Birren (1974), is that Type A individuals are slow discriminators because tension level is increased when patience is needed to take in information. The results of the study by Perry (1986) show that the Type A behaviour pattern is significantly related to the incidence of traffic accidents and violations. It also appears that the impatient component of the Type A pattern is the major contributor to the deficit in driving performance.

**Conscientiousness and accident involvement**

There is an emerging consensus among personality psychologists about a standard language for describing the basic units of personality (Goldberg, 1993). This consensus is derived from convergence in empirical work pointing towards a five-factor model (Goldberg and Rosolack, 1994). An advantage of the five-factor model is that it may ultimately provide a more systematic and comprehensive approach to the study of personality and its relations to overt behaviours like driving accident involvement, and more generally, to compliance with institutional norms and rules (Arthur and Graziano, 1996). Dimensions of the five-factor model are Extraversion, Agreeableness, Conscientiousness, Neuroticism (Emotional Instability), and Openness (Intellect).

Arthur and Graziano (1996) assessed the relations between Conscientiousness and driving accident involvement. The results demonstrated a significant (inverse) link between Conscientiousness and at-fault driving accident involvement. Conscientiousness was also inversely related to moving violation tickets. This outcome is encouraging because it suggests the Conscientiousness/accident involvement relationship is replicable, fairly stable and robust. Other than Conscientiousness, the remaining five-factor model dimensions are less clearly related conceptually to driving accident involvement. There may be links between Agreeableness and accident involvement (Hansen, 1989). The relationship between Extraversion and accident involvement has been inconsistent (Elander et al., 1993).

The results of the study by Arthur and Graziano suggest that individuals who describe themselves as more self-disciplined, responsible, reliable, and dependable are less likely to be involved in driving accidents than those who describe themselves as lower on these attributes. Conscientious individuals may be especially sensitive to social responsibility norms. One could also speculate...
that conscientious individuals are better socialised, are more likely to follow social rules and laws, and pay more attention to detail (Digman and Inounye, 1986). Within this framework, it is possible that Conscientiousness (and Agreeableness) evolves from a common developmental substrate linking attention to performance standards (Ahadi and Rothbart, 1994). Persons high in Conscientiousness may be especially attentive to performance norms or may be more highly motivated to meet performance standards than persons low in Conscientiousness (Friedman et al., 1995).

More generally, Conscientiousness may moderate people’s responses to social and institutional constraints. Highly conscientious people may have developed an adaptive self-regulatory response to the frustrations that come with institutional constraints. It is possible that the self-regulatory behaviour of highly conscientious people could lead to special problems and pathology. For example, Goldberg and Rosolack (1994) show that right-wing authoritarianism is a configuration of high Conscientiousness and low Openness to Experience.

A.1.3 Culture of driving: individuals in a social environment

Each driver is an individual influenced by the social environment consisting of other road users, general social norms, traffic-related rules of conduct, and their representations. At the same time, each driver is also sharing in the collective influence on other individuals. Many safety programmes appear to depend, therefore, on two related assumptions: (i) that drivers are sensitive to the influence of others and, and (ii) that a small shift in the behaviour of few can be amplified, through the interaction between individuals and their collectives, to a larger effect, resulting in a changed social environment or a modified culture of driving (Zaidel, 1992).

However, driver behaviour modelling lacks a connection between models of individual driver behaviour and the (presumably) resultant population behaviour as reflected in traffic characteristics, driver surveys, or driver records.

Understanding the interrelationships between individual behaviour and that of a population of drivers is important in many areas of transportation, traffic control, and safety management, when the success of a programme or a control measure depends, in part, on people’s accepting or rejecting it. Traffic safety campaigns, beyond generally promoting safety awareness, often include specific behavioural objectives. They encourage traffic-related behaviours that are deemed desirable (such as use of restraints) and attempt to discourage behaviour considered to be unsafe.
Regardless of the specific content or motivational context of the culture or social climate involved, campaigns assume that attitudes, beliefs, and behaviours of a small group of drivers may influence others in the larger group. Evolvement of a new culture can be seen as the gradual shift in the relative size of groups (Zaidel, 1992).

An essentially similar philosophy underlies the enforcement of traffic laws and regulations and, indeed, the administration of justice in general. While at any given time, only a small number of individuals are actually treated by enforcement agencies and the courts, it is assumed that knowledge about individual experiences somehow propagates to create a “general deterrence effect” in the general population.

Often there appears to be little carry-over effect or change beyond the immediate direct effect of the measure upon the individual involved. For example, Helliar-Symons (1983) examined the effect of an electronic message (“following too closely, move apart”) that was flashed at drivers. However, the results of the experiment suggested only a small, site-specific effect of the flashing sign and no carry-over to other locations was evident.

On the other hand even a large-scale, prolonged, and co-ordinated intervention designed to influence a specific behaviour, such as drinking and driving, often produces a very small and exceedingly slow rate of shift in the culture of driving. It is doubly frustrating when unplanned and undesirably traffic behaviours, such as illegal parking, seem to spread rather quickly if not checked. Wright (1981) has demonstrated, with a simple model, that an increase in the rate of illegal parking can snowball rapidly to the point of collapse of an urban parking system. The same model shows that bringing the level of compliance back to a stable condition requires a large, non-linear increase in enforcement and for a longer duration than it took to destabilise the system.

Influence of the social environment

In the above examples about changes and shifts in traffic characteristics or the culture of driving, it was implicitly assumed that the observable behaviour of other drivers induces a change in the behaviour of individual drivers who then become members of that “other” collective of drivers. How can “others” affect the behaviour of an individual driver? The social sciences (for example, Lindzey and Aronson, 1985) suggest many mechanisms of social influence, for example: learning through reward and punishment; compliance with leaders or other authoritative figures; conforming to norms, laws, and regulations; acceptance of persuasive messages. The interest is not in the fact
that drivers may be slowed down by other vehicles, nor that drivers may slow down at the presence of a police patrol or a speed limit sign; rather, it is in the ways that drivers come to adopt a different speed pattern in general, under the influence of fellow drivers.

Others as a source of information

The event of a lead vehicle slowing down, may inform the driver to pay attention; it may inform the driver that there is a ditch or police radar ahead, and it may prime the driver for the appropriate response. The informative value of the behaviour of others is high as long as people share a common driving culture; an American in Paris may not notice, or may be puzzled by or misinterpret the behaviour of local drivers. At other times, observing what others do in a novel situation may give the observer the only clue to the required behaviour (Zaidel, 1992).

Communications with others

Risser (1990) points out the importance of communication between drivers that is based on informal gestures and unregulated use of vehicle signals and lights. The emergence of unofficial communication modes shows that existing formal rules are not sufficient to resolve potential conflicts. Wilde (1976) drew attention to the existence of informal road rules that sometimes conflict with formal ones. Rothe (1990) recorded the expectations which drivers and pedestrians have of each other and found that neither group acts as expected by the other.

Communication can clarify intent and explain otherwise puzzling or offensive behaviour, and thereby reduce misunderstanding, frustration, and conflicts. It can provide a more accurate picture of others and of how we are perceived by them.

Others as a reference group

The relativity of perception, judgment, and experience is well documented (for example, Parducci, 1978; Kahneman and Tversky, 1986). The social environment is a powerful context for assessing one’s own position. In the case of speed, for example, the others’ positions may be expressed in the speed of the passing cars. It is also present in the form of the legal speed limit, one’s impression of what most drivers do, one’s understanding of the tolerance of the police, and one’s knowledge of other people’s opinions concerning speeding.
Imitation of others

Imagine a narrow street with parking on one side only, in an area that is short on parking space. Someone parks the car on the opposite side of the street with the near wheels on the sidewalk, blocking just half a lane. Within a very short time, a line of cars will be parked in the same fashion. Fashion is a fitting name for this familiar occurrence, in that imitation appears to be the moving mechanism in changing fashions (Zaidel, 1992).

Schreuder (1969) reported an experiment in inducing, by way of example, use of low-beam headlights in built-up areas. The observed gradual increase, in three weeks, from 37% to 80% use of low-beam can be attributed to imitation. The two examples suggest that attributes that make a group or an individual salient, such as a behaviour that stands out or deviates in some way, increase the likelihood of their becoming imitated.

A direct demonstration of imitation was made by Echterhof (1988) in a controlled experiment. Naive drivers travelling on a street were made to follow a car driven by the experimenter. In half the trials, the car stopped for a pedestrian (who was also an experimenter) at a crosswalk and in the other half it did not. The lead car then turned into a side street. At the next block the naive driver was faced with a pedestrian attempting to cross. It turned out that, compared to a control condition, a higher percentage of naive drivers stopped when they had experienced the experimenter’s previous stopping-for-pedestrian behaviour. Similarly, fewer drivers stopped when preceded by a non-stopping model. Echterhof believes the drivers did not learn a new behaviour, but that the model singled out, made salient, and provided an affirmation of one of two response tendencies that the drivers have already had.

Traffic and driving behaviour theories are ill-equipped to deal with such issues. Even in an area relatively rich in empirical data, such as speed behaviour, few attempts have been made to mathematically model the interactive process of drivers affecting each other. For example, if asked to anticipate the course of change in drivers’ speed behaviour as a result of specific speed regulation, enforcement policy, and practice, there is little theory that can guide traffic researchers in the prediction (Zaidel, 1992). As a further consequence, it is difficult to optimise speed control policy since the number of options is too large to consider each empirically.

Outside the field of traffic safety, other disciplines have been studying social phenomena undergoing change. The behaviour of consumers, the spread and containment of contagious
diseases, changes in political climate, shifts in dietary habits, the rise and fall of a song on the hit parade, and the spread of a technological innovation may harbour processes similar to those of traffic phenomena. Examination of other research fields may suggest concepts, principles, and modelling approaches that are applicable to traffic behaviour too (Zaidel, 1992).

Social psychophysics and behavioural decision making

Ziadel (1992) also describes concepts dealing with the way a social context is perceived by an individual. This issue is relevant for modelling the influence of others since their impact is mediated by how they are perceived. The behaviour of others is an important source of information about the world and about our own position relative to others. What is it that we know about the traffic behaviour of others? How do we remember it? How do we as drivers perceive the speed of other vehicles, size gaps in traffic, or estimate the number of drivers who enter on red? How do we integrate and remember our own experiences in traffic?

Adaptation Level Theory (Helson, 1971), Range-Frequency model (Parducci, 1978), or Functional Measurement and Information Integration model (Anderson, 1970) are models that show how the felt or judged experience of each separate event varies predictably with the distribution of other similar events. The remembered frequency of the events and their relative magnitudes are integrated into an overall impression that, in turn, shapes the perception of subsequent events. Contrast and assimilation are two perceptual mechanisms shared by many models, one focusing on differences and the other on similarities between events. Extreme events may exert disproportionately large influence, but they may also be ignored if they are perceived as unique or not representative of the relevant series of experiences.

The rational decision maker has biases and uses various heuristic rules in the face of uncertainty (Kahneman et al., 1982); the actual experience value of an outcome is separate from its decision value (Kahneman and Tversky, 1986) and is affected by social comparisons. What we think other people think or might do is important input for our own decision process (Messick and Brewer, 1983). The fact that decisions are an ongoing process requires, just as in the psychophysics of judgment, modelling the progressive changes in the value or utility of the outcomes of previous decisions as these are considered in future choices.

Computer and laboratory simulation would be useful methods for beginning the research (Zaidel, 1992). The natural extension of a study like that would be to model the interactive process.
between many drivers and to explore the possible ways in which this limited system will organise itself. Another extension would be to include individual differences as parameters in the models. Speed-control policy implications of the most promising models could then be field-tested.

Attitudes, opinions and attributions

The study of attitudes, opinions and social behaviour (Lindzey and Aronson, 1985) provides more examples of factors and potential mechanisms underlying changes in collective behaviour.

The impact of deviates, however, is tempered by another effect known as the “fundamental attribution error.” It refers to the fact that actions by a majority tend to be attributed to situational factors and are expected to be responsive to situational variations. Similar actions by only a few persons would be viewed as more exceptional, reflecting personal dispositions (traits) that are assumed to cause the behaviour and are expected, therefore, to produce consistent behaviour across different situations.

People tend to play down situational factors – the forces and constraints of the environment – when interpreting the social behaviour of others and tend to attribute the behaviour to personality and motivation. At the same time, people have the opposite bias when accounting for their own actions, especially the less positive or less successful ones.

A number of pervasive attributional biases have been identified when social perceivers infer causal relations in social settings (Fiske and Taylor, 1984). For example, social perceivers appear particularly insensitive to the contributions made by personal and situational factors in the determination of behaviour. This bias has been termed the ‘fundamental attribution error’ (Ross, 1977), and it refers to the social perceiver’s tendency to attribute the behaviour of others to dispositional qualities rather than to situational factors (Ross et al., 1977). Thus, social perceivers appear to prefer dispositional explanations for behaviour even when situational factors which can account for the behaviour are available.

A closely related bias is the actor-observer effect. Jones and Nisbett (1972) argued that there is a pervasive tendency for actors to attribute their own behaviour to transient situational causes, whereas observers tend to attribute the same behaviour to more stable personal dispositions. In a review of the actor-observer effect, Watson (1982) suggests that the terms self and other may be more appropriate, as in many empirical studies there is not one person acting and another person
observing. Therefore, any effect tends to be limited to self-other differences in situational attributions. That is, self-attributions tend to be more situational than other-attributions.

One of the consistent claims of attribution theorists is that causal attributions are basic units of human cognition, and as such mediate emotional and behavioural responses to stimuli. Attribution theory has been usefully applied to achievement behaviour, stereotyping, helping behaviour and depression (Fiske and Taylor, 1984).

A number of previous studies have suggested that motorists are subject to cognitive biases when judging the behaviour of other road users. For example, Knapper and Cropley (1978) found that drivers would readily make inferences about other drivers’ motives, opinions and values without adequate information to support such inferences. They also found that drivers would behave aggressively towards others on the basis of these inferences. Clay (1987) found that drivers considered the ‘average driver’ to be more aggressive, impatient, selfish and intolerant than they were themselves.

One possibility is that findings such as these reflect the operation of attributional biases on drivers’ evaluations (Baxter et al., 1990). In particular, self-other differences (Watson, 1982) may be apparent in drivers’ evaluations of road users. That is, drivers may tend to see their own behaviour on the roads as determined by situational factors, whilst tending to see the behaviour of others as being more dispositionally based. If this is so, then they may be relatively unforgiving of the bad driving of others if they fail to consider that this might be due to the immediate pressures others are under, rather than being indicative of a general disregard for the rights of fellow road-users.

It may be that differing trends in drivers’ self-other causal attributions are a source of stress and aggression on the roads. If some drivers habitually assume that the apparently selfish or aggressive behaviour of other road users is due to enduring characteristics of those observed then they may feel more frustration and anger than they would if they were more willing to recognise that other road users, like themselves, sometimes make mistakes or are sometimes under pressure. Conversely, drivers may under-estimate the extent to which their own ‘bad’ behaviours, which they themselves might consider relatively innocent, induce resentment in other road users. Furthermore, it may be that accident investigators, already aware of the prevalence of motivated distortions in drivers’ accident reports, should consider that such reports might in any case be fundamentally unreliable as bases for the design of accident countermeasures (Baxter et al., 1990).
The personality versus situation bias in attributing the causes of observed behaviour in traffic situations is reflected in the distinction between traffic violation and a driving error (Zaidel, 1992). For example, when a vehicle is seen entering a no-entry end of a one-way street it is likely that the driver of the vehicle will realise his/her error after a short while. Other drivers, however, are more likely to label that driver as inconsiderate and aggressive; the law could view the driver as a dangerous violator or, at best, as an incompetent driver. Traffic researchers, as well as engineers and other practitioners have their biases; some focus on the personality and motivations of the violators (or error makers), while others prefer to identify the conditions giving rise to such errors (Zaidel, 1992).

How many vehicles does one have to observe for a shift to occur in one’s attribution of responsibility? How many vehicles with no parking fees displayed will you note before deciding that there is no need to pay fees? The way we view the other actors in the traffic environment, how we integrate the evidence about their behaviour, and how it changes our view of them will, ultimately, affect the nature of models used to describe changes in collective behaviour (Zaidel, 1992).

Modelling with drivers’ view in mind

Clearly, what is needed is an understanding of the factors that influence driver behaviour in the social-collective context as well as models that describe how these factors interact in determining the pattern and rate of change in traffic phenomena. Modelling the process of change requires consideration of real environments, proper weighing of influencing factors, and detailed elucidation of hypothesised interactions and mechanisms. It is the only way of making practical predictions about the course of change and how it might be altered.

Researchers are not yet ready to formulate a mathematical model of evolving, changing traffic phenomena. The culture of driving is not just what researchers measure objectively on a road, nor is it just the professional conceptualisation of the traffic system. It is also the more general view, or “world view” which lay people have of the traffic system. This view includes their intuitive understanding and evaluation of the different components of the system and how they interact (Zaidel, 1992). It is the layperson’s theory (Furnham 1988) of “how things work.” This view is derived from direct experience, from inferences while observing others, from instruction, from the media, and from talking to other people and accepting their ideas. This view is changing and evolving over time, just as situational environmental factors keep changing.
As Furnham (1988) points out, lay beliefs have behavioural consequences beyond the view and actions of an individual. For example, the view people have of the economy or about inflation, even if it does not agree with that of the "experts," is, nevertheless, collectively expressed in consumer behaviour and wage demands which, in turn, fuel or reduce inflation.

Similarly, also in modelling traffic behaviour one should recognise that road users' actual behaviour is determined by their view of the traffic system and that this view encompasses a complex structure of beliefs related to other systems. Drivers can describe their own views and what they think other people's "world view" is. What they think other people think and do may be as useful as knowing the actual behaviour.

The importance of widening the scope of our measures and monitoring drivers' views can be illustrated in the context of modelling changes in traffic violations over time. To many observers such changes signify shifts in the culture of driving. Consideration of the justice system informs everybody that compliance with a law depends on citizens' judgment that it is fair, that it benefits or costs all citizens in a similar way (distributive justice), and that the process of its application reflects procedural justice as well (Tyler et al., 1986).

The importance of drivers' views is not limited to the influence of enforcement or safety campaigns. The efficacy of engineering designs, administrative procedures, or technical devices aimed at improving safety may also depend, in part, on drivers' acceptance and willingness to utilise them as intended. Drivers and other road users make judgments (right or wrong) about the soundness of devices or procedures and have their own ideas about what behaviour is warranted by them. For example, drivers' earlier resistance to using safety belts reflected, in part, their lack of trust in the possibility that a fabric strap would protect them in case of a serious crash (Zaidel, 1992). Concurrently, they held beliefs about physics or bio-mechanisms that allowed for feats of force resistance to moderate crashes.

Lay beliefs about safety belts, about the influence of alcohol, about reliability and crash-worthiness of vehicles and, indeed, about the very importance of safety considerations in transportation systems evolve over time. The changes come about through mechanisms similar to those that modify driving behaviour and traffic characteristics – influence of individual experiences and learning, expert persuasion, media information, group interaction, and collective phenomena such as imitation.
Thus, the shifts in the cultures of driving consist of changes in a network of related beliefs as well as in observable driving behaviours. The task of modelling these changes and relating them to each other would certainly be a challenging endeavour. It will surely pay off in a better understanding of changes in the culture of driving and, eventually in a better ability to influence their course (Zaidel, 1992).

### A.1.4 The effects of enforcement on compliance: perceived and objective risk

The police goals of increasing apprehension of traffic law violators and increasing compliance with these laws are generally viewed as being almost contradictory (Shinar, 1985). This is because for increased apprehension, it is best for the police to reduce their perceived presence, and, thus reduce the driver’s perceived risk of apprehension. Whereas to increase compliance, it is best to increase the perceived risk of apprehension, which is typically brought about by increasing the police conspicuity and visibility. The major determinant in the compliance behaviour is not necessarily what the police do, but rather, how it is perceived by the road users. Increases in compliance can be maintained or improved without concomitant increases in the police presence on the road, and the rules that govern this effect seem to be consistent with the theory of partial reinforcement in operant conditioning (Shinar, 1985).

**Perceived risk of apprehension**

The effectiveness of enforcement in obtaining compliance with the law depends greatly on each driver’s perceived risk of apprehension. All other things being equal, the greater the driver’s perceived risk the more likely it is that the driver will comply with traffic laws.

It is possible to divide the factors that influence perceived risk of apprehension into two variables (Shinar, 1985):

1. **Threat** – The degree to which visible enforcement units are seen as representing a threat of apprehension by the driver. To have an effect upon a driver’s behaviour, an individual enforcement unit must be seen as a threat. For an enforcement unit to appear threatening, it must have the following characteristics:
   a) It must be visible – Approaching motorists must see the unit if they are to perceive a risk of apprehension. While hidden units may increase the objective risk of apprehension, they do not affect the perception of it.
b) It must be identifiable – To influence the risk of apprehension, it must be seen as an enforcement unit; it cannot be mistaken for a disabled vehicle or simply someone parked by the roadside.

c) It must be prepared – The enforcement unit must look like it is prepared to enforce the law. A driver will not perceive a risk of apprehension if the vehicle is positioned where it cannot give chase or if the officer seems oblivious to passing vehicles.

2. Density – The number of enforcement units per mile of driver travel. The greater the number of units over any stretch of a highway, the greater will be the number of drivers affected and, therefore, the greater will be the overall perceived risk of apprehension. Also, the greater the number of units, the longer each driver will be affected so, with a sufficient density, a driver may be affected over a whole stretch of a roadway.

Level of perceived risk

A driver’s perceived risk of apprehension would appear to be a product of threat and density. The risk of apprehension continues to be perceived at locations of enforcement after enforcement at those locations has ceased. Varying locations of enforcement in an unpredictable manner capitalises on this time generalisation, leaving drivers uncertain as to where enforcement is occurring at any moment and extending the range of perceived risk.

Objective risk versus perceived risk

Perceived risk of apprehension may be at considerable variance from objective risk (Shinar, 1985). Whereas perceived risk is associated with increased compliance, objective risk is associated with increased apprehension for failure to comply. The difference between the two variables arises largely in the role played by threat. If threat increases the perceived risk and compliance, it will thereby reduce failure to comply and, consequently, reduce the objective risk of apprehension. Thus, efforts to increase objective risk often seek to reduce apparent risk so as not to discourage those who are violating the law from continuing to do so. This is done primarily by reducing the visibility of enforcement units (hiding them) or making them unidentifiable (such as unmarked vehicles).

The distinction between objective and perceived risk applies only to short-term effects. Ultimately, with a relatively constant population of drivers, increased objective risk, resulting in the issuance of an increased number of citations, should begin to have an effect upon drivers’ perceived risk, with
the result that compliance eventually improves. The intervening variable that mediates this effect is the information flow that results from word-of-mouth exchanges among the drivers.

Role of sanctions

Perceived risk of apprehension is just one potential deterrent to unlawful driving behaviour. Another factor is the perceived severity of sanctions imposed when apprehended.

A.1.5 Can traffic law enforcement permanently reduce the number of accidents?

The use of a game-theoretic model (Bjornskau and Elvik, 1992) might be seen as an elaboration of the argument put forward by the various theories of behavioural adaptation (OECD, 1990). The crux of the argument in these theories is that road users are not passive (parametric) receivers of safety measures. They are strategic actors who take advantage of whatever measures available to attain their goals, which is not always to maximise traffic safety. Just as the enforcer can make drivers indifferent between observing and violating the speed limit, the drivers can jointly make the enforcer indifferent between enforcing and not enforcing.

The minimal nature of road user adaptation to enforcement

A number of studies show that the effects of police enforcement are transient and local in nature. Road users adapt to changes in the perceived probability of encountering enforcement. If enforcement is encountered, violation rates drop momentarily, but return to their previous level once the site of enforcement has been passed or as time passes since the last enforcement was encountered.

A Finnish study detected an effect on speed level of a parked, marked police car up to 1.9 km from the car (Syvänen, 1976). A Canadian study (Hauer et al., 1982) concluded that the effect of a parked, uniformed police car on speed level was halved each 0.9 km passing cars were removed from the police car. No effect on speed level could be detected beyond about 2.4 km from the police car. The effect of one enforcement period lasted for about three days. Repeated enforcement extended this to about six days. Repeated enforcement did not seem to affect drivers who passed the site daily more than those who passed the site occasionally.
Another study (Hool et al., 1983) found that the effect on speed level of a stationary, marked police vehicle lasted up to about 8 km downstream from the vehicle. Upstream, the effect could be detected about 1 km from the parked vehicle. In an experiment by Brackett and Edwards (1977), a marked police car was moved from site to site in a random fashion, in order to give an impression of massive enforcement. This pattern of enforcement extended the effects up to about 20 km from each site where the vehicle was parked.

A Norwegian study (Ostvik, 1989a) found an effect on speed level of a stationary radar up to about 2.0-2.5 km from the radar site. No effect on speed level could be detected once the radar was removed.

A Swedish study (Engdahl and Nilsson, 1983b) tried to find out how long the effects of enforcement last. Several techniques of enforcement were compared. It was concluded that drivers who pass an enforcement site reduce their speed when passing the same site for a period of about 14 days. The effects did not last any longer among drivers who had passed two, three, or four enforcement sites. Enforcement conducted by means of unmarked cars did not produce any lasting effects.

Furthermore, Ross (1982) gives several examples of the transient effects of legislation and enforcement introduced to reduce drunken driving. In the United States, seat belt enforcement campaigns have only given a temporary rise in wearing rates (see, for example, Transportation Research Board, 1989).

**The effect of stricter penalties**

A number of studies show that the imposition of stricter penalties for road traffic law violations does not bring down violation rates. This does not imply that all penalties can be removed, just that their severity is less important for their deterrent effect than their existence. In 1982, the standard fines for speeding were doubled in Sweden. No changes in speeding behaviour could be detected as a result of the new fines (Nilsson and Åberg, 1986; Åberg et al., 1989). In 1987 the standard fines were once again raised. Once again, no changes in speeding behaviour could be detected (Andersson, 1989).

Not even the introduction of jail sentences instead of fines will necessarily reduce violation rates. In the state of Tennessee, a mandatory jail sentence of two days was imposed for drunken driving.
A before-and-after study, in which the states of Alabama and Kentucky formed the control group, concluded that no lasting effect on the amount of drunken driving or the number of accidents could be detected. There was an instantaneous drop of 11% in the recidivism rate. Within one year, this effect had disappeared (Jones et al., 1988).

It is not surprising that road users tend to abide by the law if the police are observed, and violate if there are no police around. Police priorities generally follow the assumptions that there are good reasons to reduce enforcement when road users abide by the law, and vice versa. Violations and road accidents can be permanently reduced if enforcement is not reduced once it is successful (Bjornskau and Elvik, 1992).

A.1.6 An evaluation of the effectiveness of police intervention on speeding

Inappropriate speed is said to be the main direct contributor to road accident occurrence and road accident severity, in terms of serious injury and fatality (for example, Johnston et al., 1980), with other major factors, such as alcohol intoxication and risk taking, often contributing to driving at inappropriate speeds for the traffic conditions. For example, the 1991 AA study (Rolls et al., 1991) on younger drivers found that speeding errors on a set route accounted for 90% of all dangerous errors, where dangerous errors were defined “driver error involving particular liability or exposure to harm”. In addition, the AA study found that the number of dangerous errors was highly correlated with reported accident frequency, and Reason et al. (1990) found that self-reported propensity to commit driving violations was predictive of accident rate. It emerged in the AA study that the majority of drivers were not incapable of driving safely, but simply did not choose to do so, or were not aware of the extent to which their behaviour put them at risk of a serious accident. Posting speed limits (Elvik et al., 1989; Summala, 1985), posted numbers of cars detected speeding (Roque and Roberts, 1989) and police surveillance (for example, OECD, 1974; Shinar and Stiebel, 1986; Bjornskau and Elvik, 1992) all do appear to reduce the proportion of vehicles travelling at high speeds, without preventing speeding completely.

General attitude towards exceeding posted speed limits, and beliefs about the consequences of such speeding, are important determinants of intentions to speed and actual speeding behaviour (Kimura, 1993; Parker et al., 1992a), as are social facilitation effects based on comparison of one’s own speed with others on the road (Connolly and Åberg, 1993). It has been shown that speeding is still perceived as socially acceptable, with people believing that most others speed; that their peers would approve of their behaviour; and that it is unlikely to result in prosecution or an accident.
Indeed, Manstead et al. (1992) found that the greater the amount by which respondents admitted breaking posted speed limits, the greater drivers' estimates of the percentage of other road users who regularly engage in similar behaviours, and the more they overestimated the actual frequency of others speeding.

The immediate determinant of a range of volitional behaviours is intention (Fishbein and Ajzen, 1975, 1980). According to this model, a person's behavioural intention, and hence his/her behaviour, is determined by the person's attitude towards the behaviour and by the subjective norm. The attitude towards the behaviour refers exclusively to the person's evaluation of that behaviour if it were performed by himself/herself. The subjective norm is defined as the subjectively perceived influence of the social environment on the subject's behaviour. The person's attitude towards the behaviour is considered a function of personal beliefs about the consequences of the behaviour weighted by his/her evaluations of these consequences. The subjective norm is considered to be a function of personal beliefs about what referents expect him/her to do, weighted by his/her motivation to comply with these referents.

Application of this model to speed behaviour on undivided highways (Vogel and Rothengatter, 1984) demonstrates that the model's variables, i.e. the attitude and subjective norm towards speeding, can indeed predict speed choice to a considerable extent. Considering the beliefs and evaluations that make up the attitude towards speeding, it appears that four motivational factors can be identified. On the basis of the item content of these factors, they were labelled 'pleasure in driving', 'traffic risks', 'driving time', and 'expenses'. The factor 'pleasure in driving' appears to have the highest relative contribution to the driver's attitude towards speeding, the factors 'driving time' and 'expenses' the lowest. Speeders and non-speeders differed significantly on all four factors, but for 'pleasure' and 'risk', these differences can be mainly attributed to differences in beliefs about the likelihood of pleasurable or risky consequences, whereas in relation to travel time and expenses these differences were found in relation to the evaluation of the consequences. In other words, fast drivers believe more strongly than slower drivers that it is more enjoyable and less tiresome to drive fast, whereas drivers who keep to the speed limit believe that it is likely to be more risky to drive fast. Both groups of drivers evaluate risk and pleasure similarly. Thus it cannot be concluded that fast drivers are unconcerned about risks, they simply do not believe driving fast increases their risk. Both speeding and non-speeding drivers are convinced that fast driving will increase travel costs and decrease travel time. However, fast drivers evaluate travel costs less negatively and reduction in travel time more positively than others, having a higher value of time.
Separate analysis in terms of vehicle characteristics demonstrated that classification of the drivers in terms of vehicle top speed (or vehicle mass) does not yield significant differences on the risk factor, but does yield significant differences in terms of pleasure in driving. Drivers of high performance cars expect more pleasure in driving fast, but do not have different expectations about safety consequences than other drivers. In the experiment by Vogel and Rothengatter, Alfa Romeo drivers behaved in accordance with the model in that they scored highest on the pleasure factor and lowest on the risk factor, with an accordingly high level of observed speed. Saab drivers were an exception to this rule: they actually scored highest on the risk factor, that is, they consider negative safety consequences more likely when driving fast but, at the same time, they are observed driving faster than drivers of any other make; Alfa Romeo drivers included. This finding contradicts Evans' (1985) postulation that drivers take into account the possible differences in the consequences of an accident when selecting their speed, dependent on the crashworthiness of their car. It is also in contradiction with the common notion that drivers of cars with a high mass and a high level of safety features compensate for these factors by driving faster, because a lower than average level of risk is perceived in this group of drivers. In fact, the risks involved in speeding are perceived by these drivers as higher than average. A simpler explanation of these findings would be that different people buy cars for different reasons (Rothengatter, 1988b).

In a widely applied model of how attitudes influence behaviours (the theory of planned behaviour: Ajzen, 1991), intentions are given a mediating role between other cognitive factors and behaviours. Other influences upon behaviours are held to have their impact via producing changes in these proximal determinants of behaviour. This model has been applied to a number of driving behaviours by several authors. For example, Parker et al. (1992a,b) were able to predict the intentions of individuals to speed in an imaginary scenario based upon knowledge of the individuals beliefs about the likely consequences of the speeding, perceived social pressure on the individuals to speed and perceptions of control over the behaviour.

An experiment by Holland and Conner (1996) builds upon the methodology employed in the mentioned studies to investigate the effects of police intervention to speed. Results showed that the intervention worked reliably for both categories of speeders (high and low), in terms of reducing the percentage of drivers breaking the speed limit and that the effect of the intervention lasted into the week following the removal of police warning signs that is, two weeks after police presence. Six weeks later, there still seemed to be some persisting effect.
The interaction effect between type of speeder and intervention week did seem to suggest that the effect of the intervention was greater in terms of reducing the number of people breaking the speed limit by a small amount, and having less of an effect on the small number of people who broke the limit by 32 km/h or more. This fits in with data from the intention measurements which indicated that those who admitted to breaking the speed limits by a large amount in the past showed more intention to speed in the future than did those who admitted to speeding by smaller amounts.

The questionnaire measures also showed that young people and men (so young men in particular) showed greater intention to speed generally, and that this increased for young men, rather than decreased or remained the same after the intervention, as it did for other age and sex groups. Young women and older men who generally broke the speed limit by small amounts were most affected by the intervention, their intention to speed being reduced, the intention of low-speeding older women being very low to start with, and changing little. Low-speeding young men actually increased their intention to speed, suggesting that although the intervention worked well in general in terms of reducing the numbers of people who broke the speed limits, the effect on one vulnerable group, young men, was questionable. Obviously, this unexpected result would need to be replicated before drawing conclusions on the long term educational effectiveness of such interventions, but it is very clear that age and sex groupings, and groupings according to pre-intervention intentions and behaviour, need to be considered separately in such studies. It is clear that those of any age who were high speeders before the intervention had changed one important belief after the intervention, in that they were now less likely to believe that they would not be caught.

The analyses incorporating traffic flow as a covariate supported the expectation that a proportion of speeding is due to traffic density, people being more likely to break speed limits when traffic density was low. In regression analyses, the proportion of variance in numbers of people speeding (driving more than 8 km/h faster than the posted 64 km/h) accounted for by traffic flow, was approximately 18%. Importantly, although traffic slowing had a significant effect in all the analyses, it did not account for observed differences between weeks, which may therefore be attributed to the intervention. It also did not account for differences between the two directions of traffic, which may therefore be attributed to the differential effects of coming from a high speed area as opposed to a low speed area.

Research on people’s inability to set adequate safety margins (for example, Brehmer and Brendt, 1989) suggests that perception of speed itself may be faulty. Indeed, people of all ages tend to
underestimate how long it will take a moving vehicle to reach a set point, and this inaccuracy increases with increasing age over 65 (Holland, 1993). The relative inaccuracy also seems to increase with increasing speed of vehicle. Ability to judge one's own speed was of specific interest in a study on the issue of breaking speed limits, and it has been suggested (for example, Casey and Lund, 1992) that a driver adapts to a high speed such that a slower speed seems slower than it really is after travelling at a high speed for some distance. The difference in numbers of speeders between the two directions of traffic in the study by Holland and Conner is therefore important in this debate. However, the finding that the difference between the two directions was greater when considering traffic breaking the speed limit by a small amount as compared with traffic breaking the speed limit by 32 km/h or more, suggests that, in addition to a possible speed adaptation problem many drivers had simply not adjusted their speed at all, implying that attitude towards speeding was also contributing to the difference.

In addition, the study demonstrates that warning signs, even in the absence of any recent police enforcement, do have a significant impact. It is suggested that many such periods of warning or enforcement, or even continuous warnings and possible enforcement, in the shape of, for example, speed cameras, may change long term behaviour and generalise to long term attitudes towards speeding and intentions to speed in other situations, even in the absence of visible enforcement likelihood (Holland and Conner, 1996).

A.1.7 Causes and manifestations of aggression in car driving

Aggressive behaviour among motorists has become a common social problem in recent years (for example, Marsh and Collet, 1986; Mc Donald and Wooten, 1988). Chase and Mills (1973) conducted a field study to investigate the effects of driver status and sex of driver upon aggressive behaviour (sounding the horn). They found that drivers hooted more readily at high than at low status vehicles. Male and female drivers did not differ significantly with respect to such behaviour.

Hauber (1980) observed the reactions of drivers when a pedestrian crossed in front of an approaching car at a pedestrian crossing at an intersection without traffic lights. Failure to stop for the pedestrian, angry gestures or language and horn sounding were classified as aggressive behaviour. Overall, 25% of the drivers displayed such behaviour. Sex and age of the drivers were important variables: of the younger men, 33% were aggressive, but of the older women only 19%. The sex of the pedestrian was also significant: male pedestrians aroused more aggression than did
females. Aggression was greater in the afternoon than in the mornings and drivers of commercial or business vehicles were more aggressive than drivers of private cars were.

Wilson and Greensmith (1983) studied drivers, classifying them by sex, accident history, and driving exposure on a route in an instrumented car. Run time, speed changes, fine and coarse steering wheel reversals, accelerator and brake applications, lateral accelerations, gear changes, free-flow speed, signals, and overtaking were all recorded. Aggression and anxiety tests were also administered and discriminate analyses were performed on the results. In general, the personality variables did not discriminate driver groupings. With regard to driving variables, however, significant discriminations were made. Subjects who had an accident record drove more quickly on clear roads, moved about more, and overtook in traffic. When moderate-exposed subjects were examined, both female and male accident-free subjects were typified by a relatively low mean free-flow speed, infrequent overtaking, and frequently being overtaken. In addition, accident-free males did not manoeuvre about in traffic. Overall, high exposure male subjects tended to drive relatively quickly but those who were accident free were typified by adjusting their car to changing conditions.

The personality factors associated with high-risk drivers were studied by Donovan et al. (1983). They reviewed literature concerned with five broad categories of psycho-social variables contributing to the risk of traffic accidents: demographic characteristics, excessive alcohol use, personality traits, acute rates of emotional distress, and driving-related attitudes. A theoretical cognitive/behavioural model was presented to integrate the results with the influence of these different factors. The studies reviewed indicated that the use of alcohol and the presence of certain personality factors, both longstanding and acute, independently increased the probability of traffic violations and accidents. Lack of coping skills, drinking behaviour, personality traits, acute emotional stress, and driving-related attitudes are all interrelated to high-risk driving. The need for further research – especially into the link between alcohol use and hostility in older drivers, and the influence of quantity and frequency of drinking on hostility, as well as the assessment of hostility – was discussed.

An effort to differentiate between good and bad drivers was made by Roy and Choudhary (1985). The results of standardised tests and scales revealed that drivers in the accident group scored high on aggression and alcohol abuse and low on extraversion personality dimensions. In addition, the accident group showed slower reaction times and reduced glare recovery than did the non-accident group.
The influence of temperature on drivers was studied by Kendrick and MacFarlane (1986). Research was carried out during the spring and summer in Phoenix, Arizona. Perceived temperature (on temperature-humidity discomfort index) ranged from 31°C to 47°C. The results indicated a direct linear increase in horn sounding with rising temperature. More defined results were obtained by examining only three subjects who had their windows wound down and presumably did not have air conditioning in operation. It was concluded that increasing temperatures were associated with increased instrumental behaviour designed to remove a source of frustration.

Although a positive correlation between the loudness and annoyance of noises has been demonstrated (Rylander et al., 1986), the relationship does not explain why the evaluations of a given sound may vary so much among people. Another study by Shimai et al. (1990) was done to assess which environmental sounds are pleasant or unpleasant. The results showed that (i) some natural and musical sounds cause pleasant affects in most people, (ii) sounds of alarm, excretion, and scratching have unpleasant associations for most people, and (iii) some environmental sounds may be estimated to be pleasant for some and unpleasant for others.

The car was seen as a potential means of expressing anger, aggression and frustration by Marsh and Collet (1987). Zimbardo (1969) suggested that anonymity is the central “input” variable to a deindividuated state, and ultimately, to subsequent aggression. An individual becomes anonymous when he/she cannot be identified by others, and therefore cannot be evaluated, criticised, judged or punished. His hypothesis has received much empirical support (for example, Deiner et al., 1976; Prentice-Dunn and Rogers, 1980; Zimbardo, 1969, 1975). Anonymity facilitates aggression on the highway. Specifically, an enclosed vehicle may provide a sense of anonymity to its occupants. The study by Ellison et al. (1995) represented an attempt to examine the role that anonymity plays in aggressive driving behaviour. The primary hypothesis was that anonymity is positively related to aggression, operationalised as horn honking. Consistent with the hypotheses, subjects driving convertibles or 4x4s with the tops closed (i.e. anonymous condition) displayed shorter horn-honking latencies, longer honk durations and more frequent honks than did subjects driving the same types of vehicles with the tops down. In general, the results of this study are consistent with previous findings that anonymity facilitates aggression (Deiner et al., 1976; Jorgenson and Dukes, 1976; Rehm et al., 1987). It was suggested that young men made more use of the car as a weapon when it was their personally owned ‘territory’.
Sex differences were noted by Farrow and Brissing (1990). They examined gender differences as related to driving behaviour. Females more commonly had difficulty with parents than males and used drugs and alcohol more frequently. Boys more often attended speed competitions, had legal problems and dated at a younger age. Males perceived greater driving skills were demonstrated in risky situations and therefore used the car to enhance self-efficacy more than females. Male drivers also scored higher on anger/hostility and sensation-seeking scales. Even though female subjects used more alcohol/drugs and more frequently came from disturbed family backgrounds, their attitudes and behaviour with respect to DWI (Driving While Intoxicated) appeared more socially acceptable.

A cluster analysis of responses from 802 female and 724 male undergraduates to 53 potentially angering driving-related situations yielded a 33-item driving anger scale with six reliable subscales which involved: hostile gestures, illegal driving, police presence, slow driving, discourtesy, and traffic obstructions (Deffenbacher et al., 1994). Subscales all correlated positively, suggesting a general dimension of driving anger as well as anger related to specific driving-related situations. Men were more angered by police presence and slow driving, while women were more angered by illegal behaviour and traffic obstructions – but differences compensated, so there were no gender differences on total scores.

Aggressiveness had also been identified as a possible contributor to driving risk. Several researchers have found aggressiveness to be related to risky driving and accidents (Donovan et al., 1988). Also, Hemenway and Solnick (1993) studied aggressiveness specifically towards other motorists, and found it to be related to a variety of types of self-reported reckless driving. However, when examining the relation between aggressiveness and risky driving behaviour, researchers have not typically distinguished between the ‘state’ and the ‘trait’ of aggressiveness (Arnett et al., 1997). In terms of the state of aggressiveness, it is possible that people (of varying levels of the trait of aggressiveness) generally take more risks if driving when their feelings of aggressiveness are heightened, for example, when experiencing the emotional state of anger. In terms of the trait of aggressiveness, it is possible that people who are generally more aggressive than others also take more risks when driving, regardless of their mood on a given occasion. It may be that both state and trait aggressiveness are related to risky driving.

The study by Arnett et al. (1997) supports the findings of previous studies on the relation between trait aggressiveness and reckless driving, but adds the finding that ‘state’ aggressiveness – i.e. the condition of being in an angry mood – is related to episodes of high-speed driving. Thus, the
findings suggest not only that more aggressive adolescents tend to drive more recklessly, but also that adolescents drive more recklessly when they are in an angry mood. Furthermore, the findings suggest that trait aggressiveness may explain the relationship between gender and reckless driving better than state aggressiveness – males reported higher rates of trait aggressiveness and higher rates of reckless driving than females, but not higher rates of state aggressiveness.

**Summary of research**

Most research has been committed to understanding the difference between drivers who are aggressive and those who are not. Among aggressive driving causes or associated features are (Lowenstein, 1997):

- a) Type A personalities (externalise anger).
- b) Life stresses such as having a cause for anger due to other factors at home with wife, children, in the work situation or frustration in general.
- c) Being easily irritated by other drivers.
- d) Tending to dehumanise other drivers, i.e. showing lack of understanding for their driving.
- e) Feeling safe, within the car environment, to express personal anger and aggression.
- f) Self-deceived status and sense of importance.
- g) The tendency to express anger outward rather than inward.
- h) Habitual responses based on habits formed over a period of time.
- i) Driver-related attitudes leading to behaviour, for example, being easily frustrated under certain circumstances such as being overtaken or not being able to overtake.
- j) Those who honk their horns are more likely to be more aggressive and there is more honking at high than low status vehicles.
- k) Men are especially likely to be lacking in patience and show aggression towards slow driving whereas women are more angered by illegal behaviour.
- l) There is likely to be more car honking and feelings of aggression when the temperature or climate is high. There is also a positive correlation between the loudness and annoyance of noises.

**A.1.8 Sensation seeking and risky driving**

Ever since Tillman and Hobbs (1949) stated that "a man drives as he lives", there has been interest in the driver’s personality as an underlying causal factor in driver behaviour. This interest has waxed and waned through the intervening decades with periodic reviews to punctuate the present
state of knowledge (for example, Adams, 1970; Beirness, 1993; Donovan et al., 1983; Signori and Bowman, 1974). However, Wilde (1994) has contented that personality has a little role to play in collision involvement and that where relationships are found, they are generally weak and inconsistent.

Nonetheless, there is a considerable body of literature which examines the relationship between the personality construct of sensation seeking (SS) and risky driving.

Nature and measurement of sensation seeking

According to Zuckerman (1994), sensation seeking (SS) is a trait defined by the seeking of varied, novel, complex and intense sensations and experiences and the willingness to take physical, social, legal and financial risks for the sake of such experiences. Key to this trait is the optimistic tendency to approach novel stimuli and explore the environment. Zuckerman (1994) views SS as part of a broader trait referred to as “impulsive sensation seeking” which is closely related to, if not part of, Eysenck’s (1983) Psychoticism dimension. SS correlates moderately with impulsivity and weakly with extraversion (Zuckerman, 1994).

Sensation is operationally defined in terms of scores on the Sensation Seeking Scale (SSS) which was first published by Zuckerman et al. (1964). The 40 forced choice items on this scale require subjects to choose between a statement which reflects a desire for sensation (“I like wild and uninhibited parties”) and one that reflects a more cautious predilection (“I prefer quiet parties with good conversation”). It is important to note that there are no items on the SSS which refer to driving behaviour.

Factor analyses have indicated that there are four dimensions which make up the scale:

a) Thrill and Adventure Seeking (TAS);

b) Experience Seeking (ES);

c) Boredom Susceptibility (BS); and

d) Disinhibition (Dis).

These dimensions of SS, while modestly correlated, appear to measure different components of SS which have been found to relate differently to various risky behaviours (Zuckerman, 1994).
Sensation seeking has been found to be higher in males than females, and for both males and females, SS increases with age until about age 16 and then declines with age. SSS scores also tend to increase with the level of education and occupational status of the individuals and that of their parents although the relationship is not linear. In terms of construct validity, SSS scores positively correlate with a variety of risky behaviours, including injury proneness, sexual activity, gambling, financial risk taking and smoking (Zuckerman, 1994), as well as risky driving.

One of the weaknesses of the studies on sensation seeking and risky driving is that virtually all of the studies on this topic have used the SSS (Zuckerman et al., 1978). Although the scale has been widely used and has been useful in measuring the construct of sensation seeking, it has a number of psychometric weaknesses (Arnett, 1994). First, the items in the scale are often confounded with the behaviour predicted by the scale. For example, the scale has been used to establish a relationship between sensation seeking and alcohol use and sexual behaviour, even though the SSS itself contains items on alcohol use and sexual behaviour. Second, the language on some of the items of the scale is awkward and dated (for example, 'hippies' and 'jet set'). Finally, the scale uses a forced-choice format where both choices may be equally appealing or unappealing to respondents, and this lowers the motivation of some adolescent respondents to answer the items carefully. Arnett et al. (1997) used a different measure of sensation seeking than previous studies, however, the same robust relation between sensation seeking and driving recklessness was found.

In a review of the literature on SS, Jonah (1997) focussed on drinking and driving, risky driving behaviour, including other risky behaviours such as non-use of seat belts and speeding, and, finally, the consequences of risky driving (i.e. collisions and citations for traffic violations).

**Sensation seeking and drinking and driving**

In summary, of 18 studies that have looked at SS and drinking and driving behaviour, all but five found positive relationships. As SS increased, reported DWI increased or reported/convicted DWIs offenders had higher SS scores. Few studies compared the subscales of the SSS but the Dis (Disinhibition) subscale was found to correlate most strongly with drinking and driving by Arnett (1990). Generally, the relationship between drinking and driving and SS has been stronger among men than women and may decline with age. There is some evidence that SS may influence DWI behaviour of men and women differently (i.e. more direct for men). Finally, the perceived risk of collision involvement while drinking and driving appears to mediate the relationship between SS and impaired driving.
Sensation seeking and risky driving behaviours

Heino et al. (1992) administered a Dutch version of the SSS to 103 male drivers, as well as a driving questionnaire. High SSs were more likely to report speeding on urban streets and highways than low SSs. Subjects, who were selected based on their high or low SSS scores, drove an 18 km route between two cities and back during which time their following distance was measured. When asked to select their own following distance, high SSs chose a shorter distance (1.19 sec) than low SSs (1.87 sec). Even though high SSs drove closer to the vehicle in front, they did not perceive the risk of collision to be any greater, suggesting that high and low SSs select following distances with which they feel comfortable in terms of perceived risk. However, when subjects’ perceived risk in free following (i.e. when subjects chose following distance) and prescribed vehicle following (i.e. subjects required to drive closer than their preferred distance) were compared, SS moderated the effect such that the increase in perceived risk from free to prescribed following was greater for low SSs than that for high SSs. On a measure of decreased heart rate variability, which was considered to be an indicator of mental load, low SSs showed a much greater effect of driving with short prescribed following distances such that mental load was greater for low SSs than it was for high SSs. These results suggest that low SSs perceive greater risk while driving closer to a vehicle than they are used to, requiring them to concentrate more on the driving task compared to high SSs.

Lajunen and Summala (1996) requested male Finnish soldiers to complete the SSS, as well as a Driving Behaviour Inventory (DBI), and then measured the driving performance of a sub-sample of soldiers. High SSs expressed greater skill orientation in completing the DBI ($r=0.29$) but a lower safety orientation ($r=-0.36$). Skill-oriented drivers tend to drive to satisfy their need for sensation (for example, “driving fast if necessary, managing car through slide”), whereas safety oriented drivers see driving more from a safety perspective (for example, “driving carefully, avoiding unnecessary risks”). SSS scores also correlated with maximum speed ($r=0.30$) during the driving performance and with acceleration ($r=0.27$).

Moe and Jenssen (1990) used the SSS to divide 90 Norwegian soldiers into high, medium and low SS groups. The high and low subjects completed a driving test on a snowy and icy track where their task was to drive as fast as possible while avoiding obstacles. A reward was offered to the winner in each of six groups made up of equal numbers of high and low SSs. Subjects from the high SS group drove faster and were more daring but they drove better (i.e. hit fewer obstacles).
The highs also had a more elevated assessment of their driving ability. On a questionnaire, high SSs rated engine power to be a more important characteristic of a car than did low SSs, whereas the lows were more likely to indicate that collision safety and fuel consumption were more important compared to high SSs.

Peer influences have been often invoked to explain adolescent reckless behaviour. In the study by Arnett et al. (1997) adolescents reported driving faster when with friends than when with a parent, however, they also reported driving just as fast when alone as when with friends. Thus, the presence of a parent inhibits adolescents’ reckless driving, but the presence of friends does not influence them to drive any more recklessly than they do when alone. It should be noted that most claims of peer influence are based simply on correlation between adolescents’ reports of their behaviour and their friends’ behaviour; the causal influence is then inferred.

High SSs’ estimate of the prevalence of the particular forms of risky behaviour among their peers is a good predictor of their own risky behaviour. This finding is subject to several interpretations. One is that sensation seekers are attracted to groups who share their own values and hedonistic philosophy. Another interpretation is that their peers model the risky behaviour and reinforce them for imitation of it (social learning). A third interpretation is that their estimates of the behaviour among their peers are exaggerated by assimilative projection; they justify their own behaviour by saying “everyone else is doing it too” (Horvath and Zuckerman, 1993).

Wilson and Jonah (1988) noted that the TAS (Thrill and Adventure Seeking) subscale combined with an impulsivity measure was a major predictor of risky driving.

In summary, all 15 of the studies that have investigated SS and risky driving other than drinking and driving have evidenced positive relationships, particularly for men (Jonah, 1997). In addition, high SSs perceive less danger in risky driving behaviour. Although not often studied, TAS appeared to be the subscale most strongly related to risky driving.

Sensation seeking and the consequences of risky driving

The ultimate measure of risky driving is involvement in collisions, particularly where the driver has been deemed at fault. Driver licence regulators have implemented elaborate demerit point systems assuming that drivers with multiple moving traffic violations are more likely to become involved in collisions in the future, and hence should be taken off the road and/or directed to improvement.
programmes. Hence, traffic violations are often used as an intermediate measure of risky driving between behavioural measures, observed or reported and the ultimate measure of collision involvement. Several studies have looked at the relationship between SS and these consequences.

For example, Moe and Jenssen (1993) had Norwegian driver education students complete a questionnaire which measured SS and self-reported collision involvement, as well as distance travelled. Low SSs had a higher number of kilometres travelled per violation compared to medium or high SS subjects (i.e. lower violation rate), but there was no significant difference on collision rate. The Dis subscale appeared to better discriminate among the three groups than the TAS subscale.

In summary (Jonah, 1997), of the 18 studies relating SS to the consequences of risky driving, most have reported positive relationships. Of the 11 studies examining traffic violations, six reported significant correlations, and three identified clusters which included drivers with high SSS scores and violations. Of the 12 studies including a separate measure of collision involvement, as opposed to an index which combined collisions and violations, seven studies reported significant differences between high and low SSs on collision involvement using correlations or analyses of variance, and another study using cluster analyses identified a positive relationship between SS and collisions. High SSs are more likely to experience collisions and violations than low SSs. Although few studies have looked at the SSS subscales, the TAS would appear to be most strongly related to driver records.

Summary of research

Several studies note that high SSs perceive less risk in various driving situations and the perceived risk and risky driving are negatively correlated (Arnett, 1990; Heino et al., 1992; Horvath and Zuckerman, 1993; Yu and Williford, 1993), suggesting that risk perception may mediate the relationship between SS and risky driving. High SSs may not perceive certain driving behaviours as being risky because they feel that they can speed, follow closely, or drive after drinking and still drive safely as a result of their perceived superior driving skills.

Alternatively, high SSs may initially perceive their behaviour as being risky but accept the risk in order to experience the thrill of engaging in it. Once high SSs have experienced risky driving behaviour which has not resulted in negative consequences, however, they may lower their perceived level of risk and engage in the behaviour more often in the future.
This notion is consistent with Fuller (1984) conceptualisation of driving as threat avoidance. When confronted with a discriminative stimulus (for example, approaching a sharp bend), the driver may choose to elicit an anticipatory response (for example, continue at the current speed). This decision depends on a number of factors, one of which is whether previous responses in similar situations have resulted in positive or negative outcomes. For example, if drivers have slowed down in the past at this curve and have discovered that there have never been any oncoming vehicles, they may conclude that slowing down is a waste of time, and consequently see no need to slow down next time. On the other hand, if they have occasionally seen vehicles coming the other way, they may engage in a partial anticipatory response by slowing down slightly or by focusing their attention and using both hands on the wheel. Once into the curve, it may become necessary to engage in a delayed avoidance response such as hard braking or swerving to avoid an approaching vehicle. The intense arousal associated with such an outcome would likely result in the driver slowing down next time.

Fuller speculates that there may be two predominant driving styles: anticipatory avoidance driving and delayed avoidance driving. He characterises drivers exhibiting the latter style as driving faster approaching hazards, experiencing high arousal in driving, tolerating high risk, being more extroverted and having higher collision involvement. This characterisation sounds similar to the profile of the high SS driver (Jonah, 1997).

Biological basis of sensation seeking

Given the ubiquitous relationship between SS and risky behaviour, Zuckerman (1994) and his colleagues have been studying the biological bases of SS. Zuckerman believes that monamine neurotransmitters like dopamine, norepinephrine and serotonin underlie the trait of SS. These neurotransmitters are involved in the production of the neuronal synapses which transmit messages from the brain throughout the body. Dopamine appears to motivate the exploration of the physical and social environment and provides positive arousal and reward associated with novel and intense stimulation. Similarly, norepinephrine provides the arousal associated with novel stimuli and amplifies reaction to them, while serotonin tends to inhibit behaviour in the face of potentially threatening stimuli. The enzyme monoamine oxidase (MAO) appears to regulate the production of these monoamine neurotransmitters by keeping them in balance.
Zuckerman (1994) notes that MAO, which has a strong genetic determination, has consistently been lower in high SSs than in low SSs. There is also intriguing evidence to suggest that individual differences in SS may be genetically based and that these genetic differences are reflected in the different biological makeup of high and low SSs. Eysenck (1983) has estimated, using twin study data, that as much as 70% of the reliable variance of the underlying trait of SS is genetic in origin.

Implications for prevention

The body of research on the relationships among SS, risky driving and biological differences, suggest that risky driving may be, at least to some extent, genetically predisposed. This leads one to consider the implications for road safety of such a genetic predisposition, if in fact it is borne out by further research. If high SS drivers can be reliably identified, what prevention measures could be implemented to reduce in incidence of their risky driving?

The most direct avenue for influencing these drivers is through the licensing system. If applicants for a new licence were screened through the use of the SSS, high SSs could be identified. While it seems improbable that these drivers could be denied a licence on the basis of their SSS scores alone, high SSs could be required to complete an educational programme focused on deterring them from using driving as a means of stimulation. The educational programme could highlight the consequences of risky driving and encourage high SSs to pursue sensation through other less destructive activities such as skiing, mountain biking or rock climbing. Through greater self-awareness, some high SSs might modify their risky driving tendencies. Similarly, problem drivers who come to the attention of the licensing authorities due to the accumulation of collisions and demerit points could be screened using the SSS to identify high SSs.

Undoubtedly, some high SSs will not be influenced by educational measures since these drivers actually enjoy the thrill of risky driving and avoiding negative consequences. Furthermore, enforcement programmes may fail to deter some high SSs risky driving since they enjoy the thrill of breaking the law and avoiding detection.

It is postulated that high SSs adapt their behaviour to take advantage of improvements in safety to a greater extent than low SSs and that, consequently, they exhibit greater risky driving. For example, if drivers were provided with a warning device which alerts them to the fact that they are becoming tired, it might be expected that high SSs over time, would adapt to the presence of this device more
than low SSs and hence would drive longer, drive at higher speeds, take fewer breaks and make more diving errors. Future research is being pursued to test this hypothesis (Jonah, 1997).

A.1.9 Objective and subjective risk, risk perception and driver behaviour

It is often assumed that behaviour is more directly determined by subjective estimates of risk than by objective risk. While this may be the case for events about which there is little experience such as nuclear accidents, it is certainly not the case for everyday events such as the risks encountered on the road (Howarth, 1988).

These relationships can be explained in the same way as other discrepancies between tacit knowledge demonstrated in skilled behaviour and conscious verbally elicited knowledge. When behaviour is well practised and automatic it does not require conscious control. Under these circumstances, conscious verbal 'knowledge' may be a reflection of social stereotypes rather than having any close relationship with the tacit knowledge which is controlling behaviour. When this happens it is very difficult to change behaviour by verbal instruction or propaganda.

It follows from the argument that the most effective safety measures are likely to be those which operate directly on behaviour, rather than indirectly through manipulation of conscious estimates of risk. In studies of the relationship between conscious, verbally expressed knowledge and the tacit knowledge which guides skilled behaviour, several authors have shown that there can be a surprising degree of dissociation between the two. For example, the best sportsmen are frequently very poor coaches because they cannot verbalise the nature of their skill or even distinguish, in a non-verbal way, the differences between skilled and unskilled behaviour in other people. In society, where so many skills are taught at least partially by verbal means, this dissociation may not occur. However, where a skill is self-taught, the knowledge related to it is frequently unavailable to conscious thought or verbal expression. Furthermore, it is clearly going too far to claim that explicit knowledge or attitudes are unrelated to behaviour. In some cases, irrational behaviour can be more closely linked to verbally expressed beliefs than to objective reality (Howarth, 1988).

Factors influencing risk perception

Difficulties in understanding probabilistic processes, biased media coverage, misleading personal experiences, and the anxieties generated by life’s gambles, frequently lead individuals to deny uncertainty, misjudge risks, and maintain unwarranted confidence in judgments of fact (Slovic et
In general, human information processing is hindered by biases and limitations which affect subjective evaluation of probabilities (Krewski et al., 1987).

People tend to simplify complex and uncertain information, relying on rules of thumb and tradition to shape our perceptions. Further, there are difficulties in detecting omissions in information received, in evaluating opinions, and in detecting inconsistencies in debates about risk (Fischhoff, 1985). Despite these difficulties in assessing risk, people may utilise existing information to form strong views about risks. Such perceptions are often resistant to change: new evidence which is consistent with initial perceptions is accepted, while that which is contrary is dismissed (Slovic, 1987).

People do not perceive all lives to be of equal value, nor do they perceive all forms of death as equal. Frequently, non-fatal health impairment such as permanent brain damage is seen as more serious than death itself. In the public’s view, low probability high consequence events tend to be evaluated more in terms of consequences than probability, to the point that what is possible becomes more important than what is probable (Whyte, 1984).

Psychological determinants of perceived risk

Several studies conducted by (Lichtenstein et al., 1978; Fischhoff and MacGregor, 1983) have provided insight into the public’s perception of the frequency of lethal events. These investigations have revealed two highly consistent but systematic biases in judgments of frequency. First, individuals tend to overestimate the frequency of occurrence of low probability risks and underestimate the frequency of higher probability risks. Second, individuals tend to exaggerate the frequency of some specific risks and to underestimate the frequency of others. For example, people tend to underestimate the frequency of death due to stroke, diabetes, heart disease, and cancer, and to overestimate the frequency of death due to tornadoes, floods and botulism. Test subjects in these studies were unable to correct for these sources of bias when specifically instructed to do so.

Errors in the subjective estimates of risks frequently result from heuristic biases, which occur due to the unconscious need to make sense of new information (Slovic et al., 1982b). Heuristic are inferential rules used to make generalisations using limited information, and to reduce difficult mental tasks into simpler ones. This can lead to large and persistent biases in judgment with serious implications for risk assessment (Slovic, 1987).
Scientific experiments and casual observations indicate that individuals have difficulty thinking about and resolving risk-benefit conflicts, even in simple gambling situations (Slovic et al., 1979). In order to deal with the anxieties associated with uncertainty, individuals frequently perceive patterns in the occurrence of hazardous events. Such perceptions enhance feelings of control and the belief that it is possible to predict the occurrence of future events. This principle is embodied in the gambler’s fallacy that a long losing streak necessarily heralds a change of luck.

Decisions made under condition of uncertainty can be influenced substantively by the way in which the problem is framed. A decision frame reflects an individual’s conception of the acts, outcomes and contingencies associated with a particular choice (Tversky, 1981). The decision frame adopted by an individual for any decision is controlled partly by the formulation of the problem, and partly by the norms, habits and personal characteristics of the decision-maker. Rational choice requires that preferences between options not be reversed when frames are changed. However, due to the imperfections of human perception and decision-making, changes in perspective often reverse the relative desirability of options, usually unbeknownst to the individual.

The framing of a problem is particularly important for new and complex risk issues, for which people have no well-defined preferences. In such situations, individuals may be unfamiliar with the terms in which the issue is represented, or may hold contradictory values due to their occupancy of different roles in life. They may also experience shifting views with time and be uncertain as to which should form the basis of their decisions. Even when their own inconsistencies in judgment are recognised, they may not want to resolve the corresponding conflicts, resulting in an impasse (Fischhoff et al., 1979).

Sociological studies have demonstrated that risk perception is influenced by both social and cultural factors (Klein et al., 1993). The opinions and actions of friends, family, co-workers and respected public officials all contribute to an individual’s perception of risk. Further, certain individuals tend to consciously or unconsciously emphasise certain risks and de-emphasise others when fulfilling their roles in social groups for the purpose of maintaining or controlling the group (Slovic, 1987).

**A.1.10 Drivers’ perception of specific hazards on the road**

It has been argued (Taylor, 1964; Näätänen and Summala, 1974; Watts and Quimby, 1980; Wilde, 1981; Fuller, 1984) that road users modify their behaviour according to the risk they perceive.
Taking part in traffic is largely a self-paced task so that accident risk is under the road user's voluntary control (Janssen, 1990). If so, then it would be valuable to identify those aspects of the environment which road users take into account and how much each aspect contributes to the perceived risk. However, no suitable method for assessing road users' perception of individual hazards has yet been developed (Armsby et al., 1989). Such a method would have many benefits in the work of reducing road accidents. It would enable existing theories regarding the causal factors in road accidents at specific sites to be clarified, aid engineers in the design of remedial work and help in the development of driver training programmes.

None of the interview techniques used by Armsby et al. (1989) in their study, was successful in drawing a rich response from drivers on the subject of road hazards, which were rarely mentioned. When road hazards were mentioned, they usually had to do with the behaviour of other road users rather than with aspects of the road environment. Subjects appeared to view the road as a fixed landscape, which it was their job to negotiate successfully. Their responses were not concerned with hazardous features of the environment as such, but with the driving task given the need for coping with those features. The interviews often gravitated towards the inconvenience and frustrations of driving, with occasional comments about the hazards generated by other road users. In general, young drivers seemed to be more concerned with hazards which attracted attention than old drivers were, whereas old drivers were more concerned with predictability – they recognised that the most obvious hazards are not necessarily those which are most risky. The observed differences in the perception of parked vehicles seemed to fit into this broad pattern. Parked vehicles were perceived as part of the road environment by young drivers, but old drivers saw them as separate entities, perhaps with potential for movement, and therefore a potential source of risk.

One possible reason for the low level of responses concerned with specific 'road environment' hazards, is inadequate learning (Armsby et al., 1989). Drivers may be aware only of those hazards which experience has taught them may lead to near-misses or accidents, or which have been brought to their attention in some other way. It is possible that drivers are generally unaware of the large variations in risk from one location to another, and view the road environment not as a place at which accident risk varies systematically from place to place, but as a neutral setting in which other road users suddenly and indiscriminately threaten their safety. Studies which have compared the subjective assessments of drivers at various different types of sites with the accident statistics of those sites, have shown that drivers do not always make an accurate assessment of accident risk (Watts and Quimby, 1980).
Interestingly, car drivers who also drove (or who had driven) motorcycles were more specific about road hazards than car-only drivers were, particularly when speaking from the motorcyclist's point of view (Armsby et al., 1989). This might be expected, given that motorcyclists are more at risk from physical deficiencies in the road environment, such as a road surface with low skid resistance, and more vulnerable to injury if they are involved in an accident.

Armsby et al. conclude that the subjects' responses were idiosyncratic because they were closely related to personal experience. They were relayed in conversational language, which made extraction of the salient features difficult. In many cases, it appeared that these personal experiences were recalled by use of visual imagery.

In fact, in a study by McKeller (1965), 97% of the subjects reported by using visual imagery. Imagery is often used to remember information more efficiently (Paivio, 1971). However, the fact that subjects could not talk generally about road hazards, and appeared to rely substantially on visual imagery, suggests they only thought of hazards as specific events in the past. They found it difficult to generalise from their individual experiences of hazards to form an overview of their importance in overall terms. It would seem that fixed road features are not perceived as hazards until a particular set of circumstances, which the subject believes can lead to an accident or near miss, turns them into hazards for the individual concerned (Brown and Groeger, 1988). This is consistent with the earlier hypothesis of inadequate learning.

The identification of potential hazards must be represented internally in spatio-temporal terms (Brown and Groeger, 1988). This may partly account for Mourant and Rockwell's (1972) finding that novice drivers have a different visual fixation and scanning pattern from that of experienced drivers. Novices will also learn the dynamic characteristics of vehicles and pedestrians, which allow them to predict the trajectories and nature of hazard presented by moving objects. Other research on drivers' steering control (McLean and Hoffman, 1971) suggests that the higher-order steering cues which drivers learn to use as they gain experience help to direct their gaze farther ahead, thus increasing the probability that upcoming hazards will be identified early. There is support for this view from research reported by Brown (1982a), showing that young novice drivers are relatively poor at identifying distant traffic hazards, although they compare well with older experienced drivers in identifying near hazards.
A.1.11 Drivers' perception of risk in automobiles

One of the well-established areas in which ergonomic principles have been employed (and also highly advertised), is the design of automobiles. Although there are still some problem areas, automotive manufacturers have employed human factors principles in a variety of ways. In particular, improvements in safety devices have reduced the effects of automobile accidents substantially (Leonard and Karnes, 1998). At the same time, many people seem to think that these improvements will automatically produce safety without effort by the driver or passengers. The level of knowledge about some automotive hazards, for example tires, is relatively poor (Leonard, 1997). The physics of inertia is also difficult for many individuals to evaluate correctly, as seen in the work of McCloskey and his colleagues (McCloskey et al., 1980; McCloskey et al., 1983).

One’s ideas about movements and inertia in various situations can influence one’s actions in the automobile. Consumer concern with safety may be nullified because understanding of hazards associated with uses and misuses of products is often inadequate owing to lack of understanding relevant physical principles.

Current thinking about automotive safety emphasises the use of passive restraints. Research has strongly supported the effective impact of seat belt use on reducing or avoiding traffic injuries and traffic fatalities (Hunter et al., 1993; Evans, 1996). Despite a wealth of information about the importance of using seat belts, 25% to 35% of individuals fail to use them (Leonard and Karnes, 1998). Factors affecting the level of wearing rates include socio-demographic characteristics, such as gender (Preusser et al., 1991); age and social status (Liu et al., 1998); and behavioural patterns, such as: alcohol consumption; high risk driving (Robertson, 1996); obedience to circulation regulations and State rules and the perceived benefits and barriers of safety belt usage (Crandon et al., 1996). Researchers have also studied the influence of travelling conditions and driving time on wearing rates (McCarthy, 1986; Miller et al., 1998). Some individuals may specifically decide not to use seat belts. The reasons vary from an attitude of, “You cannot tell me what to do,” to the fear of being trapped in a vehicle after a crash. In other cases, failure to use safety belts may be attributed simply to forgetting to buckle up.

A Principal Components analysis by Chliaoutakis et al. (2000) indicates that the factors which characterised young drivers’ motivations to use seat belts are those of ‘environment’ (for example, bad weather, narrow roads, heavy traffic) and ‘imitation’, and the factors which characterised their intentions of not using seat belts are those of ‘risky behaviour’ and ‘discomfort’ (for example, claustrophobia). However, the statistical evaluation of these factors did not prove the association
between the four factors and the self-reported frequency of seat belt use. In other words, there was a divergence between strong motivations (intentions) and behaviour (reported frequency), or as noted by the European Transport Safety Council (1996), ‘the relationship between what people say they believe and what they actually do is weak in the case of seat belt use’.

A.1.12 Accident risk and perceived driving ability of young and older drivers

Risk taking or failure of skills

Young drivers are significantly over represented among all drivers involved in traffic accidents and fatalities. Furthermore, young drivers are much more likely than older drivers to be responsible for the crashes in which they are involved (Williams and Karpf, 1984).

Shinar, McDonald and Treat (1978) found that inexperienced drivers are likely to be involved in accidents due to improper directional control of their vehicles. This is supported by data which show poorer performance for less experienced drivers on tasks, such as steering and speed control (Blaauw, 1982; Kimball et al., 1971). In addition, Quimby and Watts (1982) demonstrated that although younger drivers tend to have shorter simple and choice reaction times, they also take longer than intermediate-aged drivers to perceive and respond to potentially dangerous traffic circumstances under simulated conditions.

However, Hodgdon et al. (1981) found that this overrepresentation in collisions could not be explained by more miles driven, less driving experience or poorer mechanical condition of vehicles driven by young drivers as compared with older drivers. What did account for greater collision involvement by young drivers was the manner in which they drove. Young drivers speed more frequently, drive through yellow lights more frequently, wear seat belts less frequently and accept shorter temporal gaps when entering traffic – in other words, young drivers take more risks than older drivers.

It seems reasonable to argue that young drivers have a greater risk of accident involvement because they are not as experienced and consequently, their driving skills are not as proficient than those of older, more experienced drivers are. However, the problem with this argument is that it is rather difficult to prove since driving experience and exposure to risk are often confounded (Brown, 1982a).
Finn and Bragg (1986) suggest that young drivers do perceive the risk of an accident significantly differently than do older drivers – and in the direction of perceiving less risk than do older drivers. Both groups saw young male drivers as having a greater risk of an accident among all drivers. However, young drivers saw their own chances of accident involvement as significantly lower than those of other young male drivers, while older drivers saw their own chances of accident involvement as comparable to those of their peers. Brown and Copeman (1975) report that young male drivers rated a given series of traffic offences as potentially less hazardous when they were notionally responsible for committing them, than when the offences were notionally committed by other drivers. In addition, results from Matthews and Moran (1986) show a trend for lower self-ratings in ability for drivers with a poor accident record. At the same time, drivers who have had more accidents rate the overall risk of an accident higher.

Analysis of the subjects’ ratings of videotaped sequences (Finn and Bragg, 1986) showed that the older drivers rated a tailgating situation as significantly more dangerous than did the young drivers, but a surprise pedestrian situation was rated significantly more hazardous by the young drivers. In terms of on-road behaviour, however, the tailgating discrepancy between the groups is likely to reflect frequent hazardous driving on the part of young drivers, while the pedestrian discrepancy is not likely to lead to safer driving by young drivers. Avoiding collisions with pedestrians who have yet to be spotted requires a consistently safe driving approach that anticipates danger – an approach young drivers appear to lack – while failing to see the risk of tailgating can lead to constantly placing oneself (and others) in jeopardy by failing to allow a sufficient gap between cars.

Knowledge of one’s driving capabilities and skills is dependent upon both short- and long-term memory of driving experiences. In addition to being updated by ongoing driving behaviour, such information will contain boundary conditions which, on the bases of previous risk-taking actions, help the driver to set limits upon and evaluate potential outcomes and risks of available action options (Matthews and Moran, 1986).

Taken together, the findings suggest that risk misperception – or least seeing less risk in driving situations than older drivers – may account for the disproportionate involvement of young drivers in traffic accidents.
Process of gaining experience as a driver

During the first years and the first 50,000 km of driving, two principal developments occur among inexperienced drivers (Summala, 1987b). Firstly, advanced and automated driving practices and hazard control skills develop slowly with exposure to traffic. Secondly, the car gradually loses its role as a means of satisfying extra motives such as showing off and sensation seeking. Problems arise especially from the fact that young drivers seem to achieve, quite soon, the basic car control and guidance skills which bring them overconfidence in their abilities while the hazard control skills still are defective and the extra motives are most influential. To gain experience, young drivers must spend time behind the wheel, but in doing so they increase their chances of being involved in an accident because of their inferior skills. Warren and Simpson (1976) have referred to this dilemma as the "young driver paradox".

As to basic driving skills, it is possible to lengthen driver education and make the licence examination more stringent and, if necessary, to delay licensing until basic car control skills have been acquired. Only when learners have shown in another driving test six months later that they have acquired certain traffic skills will they get a permanent licence. However, it is only a part of the problem. Advanced hazard perception and control skills require substantial time on the road (Summala, 1987b).

On the other hand, young drivers can be prevented from bringing their motives and social norms to the car. Car advertisements, other drivers, and even legislation provide young drivers with behavioural models which are far from safe. In addition, inexperienced drivers should be shown very concretely that they are not able to manage in some critical situations.

Summala (1987a) warns against too much optimism, however. From the well-known results of Williams and O’Neill (1974), which show a higher accident involvement (and a related higher speeding violation rate) among race drivers, advanced driving skills do not save drivers from accidents. Furthermore, Summala measured driving speeds of Finnish road-safety researchers when they were approaching a hotel where a national road safety meeting was to be arranged. Of the 13 road-safety researchers who could be tracked by the radar in the speed limit zone of 60 km/h, nine exceeded 70 km/h, six exceeded 80 km/h and three reached 90 km/h, which means gross endangering of traffic in the Finnish legislation. The researchers’ speeds were clearly faster than those of other traffic were. These results definitively show that neither skills nor knowledge can save one from risky driving and accidents. The stochastic nature of road accidents, automated and
habitual driving practices, and the feeling of control at the task deceive an expert as well as an inexperienced driver.

A.1.13 Driving skills: experience and expertise

After passing a driving test and obtaining a licence, the ordinary motorist is left largely to his/her own devices during the subsequent years of experience on the road. Driving involves a large set of separate skills, such as information acquisition (especially scanning) and perceptual-motor coordination, anticipation and assessment of the traffic situation, risk estimation, setting safety margins, balancing the disparate attractions of speed and caution, and so forth. The question is what happens to driving skills during this extended period of unsupervised experience. There are reasons to suspect that experience may not always be beneficial (Duncan et al., 1991a).

From research it is clear that novices drive relatively badly, and skills improve with experience. Duncan et al. found that novices were the worst group in most car control skills (late acceleration, poor and inconsistent steering path, slow gear change). Poor car control in novices was also noted by Quenault and Parker (1973) and Michiels and Schneider (1984). Poor path round a bend is specifically described by Cavallo et al. (1988). Quenault and Parker (1973), for example, compared age-matched groups with between 1 and 52 weeks of post-test experience. Car control (as rated by an experimenter) improved with experience. Compared with a random selection of older, more experienced drivers, furthermore, all these (relative) novices showed more lapses, errors and near accidents, drove more slowly and used their signals less often. Other studies have confounded experience and age, but the results are similar. Mourant and Rockwell (1972) found that, compared with an experienced group, drivers who were young and inexperienced tended on the average to look less far forward and further away from the centre of the road, perhaps looking to the kerb to judge lateral position. They were also less likely to check the driving mirror. However, when age is controlled, it has proved hard to demonstrate a clear relationship between experience and actual accident involvement (OECD, 1975).

On theoretical grounds changes with experience may be more complex than a pattern of simple improvement in all aspects of driving skill. To illustrate, Duncan et al. (1991a) consider two different errors – mistiming the components of a gear change, and failing to check the mirror before a manoeuvre. A grinding of gears immediately punishes the first, but the second will very rarely produce any perceptible effect. It seems rather likely that such differences will be important in determining the course of skill change with experience. Like any complex task, driving is a matter
of balancing disparate goals, such as speed, smoothness, and caution. If increasing experience alters the weighting of different goals, this too may make some aspects of performance less satisfactory.

Using an instrumented car driven in normal traffic, Duncan et al. assessed the driving skills of trained experts, of normal, experienced drivers, and of novices. To deal with experience uncontaminated by age, novices who had obtained their driving licence comparatively late in life, were chosen. In this group the mean age of obtaining a licence was 35, as compared to 22 for the normal, experienced drivers. During the experiment, it was often the normal, experienced drivers who performed worst, while novices sometimes resembled experts. These results support the view that simple experience on the road is not sufficient to produce improvements in all driving skills. Experience seems to be a matter of enhancing skill in some areas, but allowing bad habits to develop in others. As Schneider (1985) has emphasised, practice does not always make perfect. At least for some skills, furthermore, there may be reason to suppose that the bad habits of experienced drivers are related to accident involvement. Quenault (1967, 1968a) found satisfactory mirror checking to correlate with freedom from lapses, risks and near accidents on a test drive. Quenault (1968b) also described a wide range of errors typical of drivers whose mirror use is poor, including ignoring or seeming unaware of the traffic environment, ‘getting lost’ in the organisation of manoeuvres, and overtaking dangerously. At least in part, therefore, people who check the mirror rarely or follow too closely may be over-involved in accidents because of other errors they tend also to commit. It is possible that factors such as socio-economic status and cautiousness are related to the age of learning to drive, and therefore it is conceivable that these, rather than experience per se, are responsible for some of the differences Duncan et al. observed.

Both theoretically and practically, it is important to ask why experience improves some driving skills but not others. Duncan et al. suggest that the nature of feedback in unsupervised driving may be crucial, and on the whole results are consistent with this hypothesis. Thus, experienced drivers are relatively poor in aspects of mirror checking, anticipation (rightward looks on approach to a roundabout, braking into an intersection) and setting a safety margin, for all of which it is easy to argue that immediate feedback from the driving situation is usually weak. On the vast majority of occasions, no negative consequences will result from a failure to check the mirror or from following too closely; and as for anticipation, by definition its consequences are only felt at least a second or two after the actual behaviour. These then may be the sorts of skill most dependent on the supplementary feedback that can be provided by a tutor; and therefore the skills most likely to deteriorate once a tutor is no longer present. By contrast, experienced drivers accelerate and change
gears relatively quickly, and have efficient and consistent steering paths. Others have argued that poor progress in traffic is an especially effective immediate impetus to action, accounting in part for a bias towards speed at the expense of caution in normal driving (Summala, 1988). Similarly, it can be argued that deviations from an intended path are immediately apparent and call for immediate corrective action.

Behaviour on the road is the end product of a balance between competing goals, and that for reasons unrelated to feedback different drivers may simply be trying to achieve different goals. The balance between speed and safety is the most obvious example, but other goals include obeying the law, driving as one has been taught, and even (especially perhaps for the experts) giving an exhibition of ‘perfect’ driving skills. For example, experienced drivers brake late into an intersection simply because they prefer good progress to a smooth ride. As another example, pulling into the outer motorway lane to overtake has both benefits (avoiding close following) and costs (blocking the progress of other high-speed traffic). Novices may simply balance these considerations differently from other drivers, perhaps because they feel more at risk (Duncan et al., 1991a).

A.1.14 Theory of goal weighting in driver behaviour

In a study by Duncan et al. (1991a) a large group of normal motorists – people with at least five years’ driving experience – was compared with smaller groups of acknowledged experts and novices. The results indicated that not all of the measured driving skills improved smoothly with experience. An explanation of the results concerns feedback, or information from the environment indicating that an action is or is not important. It seems that poor feedback makes some driving goals seem unimportant. Thus, maintaining an updated record of following traffic, or an adequate distance from the car ahead, may seem unimportant, since apparently they are unrelated to either progress or safety. However, it can also be argued that in driving, as in most activities, behaviour always reflects a balance between competing goals (Duncan, 1990).

Duncan (1990) also addressed individual differences among normal drivers. Does one driver tend in general to be ‘better’ or ‘worse’ than another, or is each component skill so dependent on the specific way it has been learned as to be uncorrelated with others? On this question the findings were clear-cut (Duncan et al., 1991b). For the sample of normal drivers in the experiment a pattern of very low correlations between different component skills was reported. Even looking at groups of apparently similar skills changed the picture rather little. For example, three measures (mirror
checking at intersections, rightward looks at approach to a roundabout, rear looks before pulling in after overtaking) had to do with scanning during different manoeuvres. Their median inter-
correlation was 0.16. Four measures had to do with car control at intersections (early braking on entry, rapid acceleration on exit, absence of under-steering, consistency of path across repetitions). Their median inter-correlation was 0.14. Though the overall correlation matrix did show the usual property of positive manifold – the average correlation of each measure with all others was marginally positive – the striking result was how independent each skill was of all the others. Similarly, there were no large correlations between driving skills and any of a battery of laboratory tests (such as general intelligence, reaction time) administered in a separate session.

In the dual task study, subjects were asked to drive twice round the test route, once with a concurrent task (Duncan et al., 1991b). The task chosen was to generate a random stream of letter names out loud at short intervals. This task shares little content with driving, but is mentally demanding and is hard to automatise. In the dual task study, concurrent random letter generation produced only modest decrements in the measures of driving skill. Across the measured driving sub-skills, however, profiles of dual task decrement and general intelligence (or Spearman’s g) correlation were in excellent agreement. The correlation between profiles was 0.67. Those skills which correlate most strongly with g are also those which are most sensitive to a demanding but dissimilar concurrent task.

To summarise the results on individual differences: for particular driving skills, individual differences must depend almost entirely on very specific learning. Correlations with other measures, even other, similar driving skills, are uniformly low. Across skills, however, the profile of small correlations with ‘general intelligence’ agrees excellently with a profile of dual task decrements (Duncan, 1990).

Theory of goal weighting

In simple stimulus-response systems, the choice of behaviour is dictated only by the current stimulus or world state. Thus the basic unit of analysis is a stimulus-response pair. In a goal-directed system, behaviour is controlled not only by the current state, but also by a description of the desired or goal state. The basic unit of analysis is a triple: current state, goal state, action taken to diminish the difference between the two states.
To achieve a fixed consequence or goal state, behaviour should be controlled by a description of this goal, and operations should be chosen to diminish any difference between it and the current state. When no difference remains and a test for goal state is satisfied, its control of behaviour is relinquished. Under these circumstances, behaviour has a natural adaptability in the face of unpredictability or noise. Since all situations have an element of uncertainty, it is always possible that some action will produce a new state other than the expected goal state. Since a test for the goal state is then not satisfied, however, further actions will automatically be chosen (based on the new current state which has been created) until such satisfaction occurs (Miller et al., 1960).

How then are goals chosen for the control of current behaviour? In artificial intelligence approaches to this question (Laird et al., 1987), goal choice proceeds through some form of problem decomposition. The programme is provided externally with the basic or ‘top’ goal, which, for example, might be to solve some geometrical problem, or to build some structure of blocks. Through applying its knowledge of the problem domain, the programme determines some set of sub-goals which, when achieved together, satisfy the top goal.

There is no question that this scheme has enormous advantages, and matches a good deal of what one observes in human behaviour. On its own, though, it is insufficient to deal with the complexity of action choice in the real world. People live not in a static problem domain but in a dynamic environment, where new events can always arise to overturn current concerns (Reitman, 1965). As Anderson (1983) and others have noted, one needs to consider not only how current goals are decomposed into sub-goals, but how stimulus inputs can activate new, unrelated goals (corresponding, in the above programmes, to the external imposition of a ‘top’ goal).

These considerations suggest an account of the following sort (Duncan, 1990). The current stimulus state activates a set of possible next-states. Each of the potential next-states offers itself as a candidate new goal state, but since there are obvious limits to how many actions can be taken at once, it is necessary to select which of the suggested next-states to pursue. Candidates must be weighted by rules specifying their relative importance, and there must be some sort of competitive process selecting the candidate with the highest weight.

Very likely, different sorts of rule are reflected in different behavioural strategies (Kuhl and Kazen-Saad, 1988). People seem typically to balance and switch between such behavioural styles, from opportunism (Hayes-Roth and Hayes-Roth, 1979) to focused problem-solving.
Most work has under-emphasised the flexibility which is provided by the possibility of working forwards from the current state (opposed by working back from the goal state). In combination with the idea of competition between alternative candidate goals, working forwards allows interruptible, flexible goal choice in a dynamic world. In human cognition, flexibility may be bought at the necessary cost of some element of distractibility (Duncan, 1990).

Behaviour directed towards some goal takes the form of producing a sequence of (world and representational) states, beginning with the current state and ending with the final goal state. Each state acts as a goal, in that actions will be directed towards it (even in the face of noise) until it is achieved. To behave coherently, therefore, is to discover a path of states linking the current state to the active goal state, and this is the process known usually as problem-solving.

The current state activates a set of potential future states, presumably with varying strength, depending on how familiar, complex or certain the path of intermediate states is by which they may be reached. At the same time, the active goal state activates sets of potential relevant states or states from which the goal state can be reached. Again, the strength of activation will reflect the familiarity, complexity and certainty of the path which would link an activated potential state to the goal state. Where the two forms of activation collide, a new sub-goal is created. Thus, the mutual constraints of activation from current and from goal states tend to attract behaviour into an appropriate path. Sometimes called a solution path, it might less neutrally be termed an attraction path.

**Efficiency of goal weighting**

Distractibility may be the price paid in human cognition for flexibility; and it seems likely that such distractibility may be stronger in some psychological states or in some people than in others. More generally, there may be wide variations in how effectively coherent streams of behaviour are produced by use of appropriate rules for weighting candidate next-states (Duncan, 1990).

A striking example is the gross behavioural disorganisation which can result from major damage to the frontal lobes of the brain (Bianchi, 1922; Luria, 1966). This disorganisation can reflect itself in any behavioural domain. A key aspect is that, though a task’s requirements are understood, the corresponding behaviour may be inappropriate. Behaviour in violation of task requirements may be continued, or behaviour may cease though goals have not been achieved.
Duncan (1990) suggests the hypothesis that impaired goal-weighting lies at the root of the frontal lobe syndrome, and that variable efficiency in this process is closely related to Spearman's $g$. The efficiency of weighting determines how easily new goal states are established lying on the path to a currently active super-goal (such as satisfying the experimenter); and when the process is seriously impaired, behaviour loses much of its goal-directed appearance. In frontal lobe patients this can be reflected with especial clarity in disorganisation of speech, which wanders continually from the point and is filled with irrelevant associations; but much the same character can be seen in all manner of behavioural domains, ranging from simple perceptual-motor tasks to complex problem-solving (Luria, 1966). Even when a ‘goal’ has been clearly described in verbal instructions, furthermore, it may not be selected for the control of behaviour – the phenomenon of knowledge-behaviour mismatch.

Though the evidence is in fact rather scanty, it has often been suggested that practice (i) renders performance relatively insensitive to frontal lobe damage (Walsh, 1978), (ii) reduces correlations with $g$ (Ackerman, 1988), and (iii) diminishes interference between concurrent tasks (Schneider and Shiffrin, 1977). These effects of practice provide the strongest reason to believe that there may be a close link between frontal lobe function, $g$, and divided attention. Commonly it is said that practice diminishes reliance on conscious, ‘executive’ control, and renders behaviour ‘automatic’. Thus, practice makes behaviour relatively insensitive to the efficiency of goal weighting.

One approach (Anderson, 1983; Rasmussen, 1986) is to suggest that well-practised behaviour is not goal-directed at all. The basic idea is that, after sufficient practice, stimulus inputs or current states directly call forth their associated actions, without reference to any goal state. Control by current state-goal state-action triples is relinquished for control by stimulus-response pairs. This seems consistent with the fact that well-practised reactions to some environmental input often appear as absent-minded slips or sources of distraction even when current goals make them inappropriate (Reason, 1979; Stroop, 1935).

Duncan (1990) suggests a different approach to these findings. A major merit of goal-directed behaviour is its flexibility in the face of noise. Even if the outcome of any given action is not entirely predictable; still, providing that behaviour continues to be controlled by a description of the same goal, actions will continue to be taken until this goal is achieved. Even in the best-practised behaviour, there is sufficient uncertainty in the environment (and/or the cognitive system) to require such flexibility, and that control by a current state-goal state-action triple is always retained.
How does one deal with evidence that well-practised behaviour is under direct stimulus control? Duncan (1990) proposes that any stimulus state activates potential next-states, and that familiarity is one major factor determining the strength of this activation. Thus, the proposed solution is that familiar stimulus states are linked not just to associated actions, but also to associated next goals states. The stimulus controls behaviour, but by activating a whole associated control triple (current state-goal state-action), not a simple stimulus-response pair. Phrased differently, in well-practised behaviour each current state provides such strong activation to the familiar next-state as almost to ensure that it will gain control of behaviour.

This account makes it clear why practised behaviour should be relatively insensitive to the efficiency of goal weighting. With increasing experience, each familiar next-state so easily wins the competition for behavioural control that the efficiency of rejecting competitors is unimportant. For a familiar problem or type of problem, any current state is already linked to next-states which have proved useful in the past, and it may make sense to proceed directly to them. For an unfamiliar problem proceeding directly from the current state makes much less sense. By definition, familiar paths leading from the current state are unlikely to approach the goal. On the other hand, paths from some possible states to the goal state will probably be obvious or familiar, even though a path from the current state is not. If familiar paths leading from the current state are irrelevant, the best strategy may be to activate next-states from which the goal state can be achieved, in the hope that one of them, when selected as a sub-goal, will in turn suggest a path from the current state.

Driver skills

Duncan suggests that certain goals (skills) – for example maintaining an updated record of following traffic, or preserving a wide safety margin – are not well supported by feedback obtained in the course of normal driving (Duncan et al., 1991a). Taking mirror checking as an example: during training, inadequate mirror checking is persistently punished by criticisms from the tutor. The result is development of a strong attraction path, linking onset of a manoeuvre (a current state) to avoidance of criticism (a goal state) through careful mirror checking (a sub-goal). When the tutor is no longer present, however, occasional failures to check the mirror cease to bring adverse consequences. Given competition from alternative next-states, for example, those requiring other fixation patterns, use of the mirror may gradually decline. The question remains whether declines in performance can be prevented by training programmes which originally are more effective in ‘stamping-in’ the desired behaviour – or whether a better approach would be to ‘top-up’ training
with occasional follow-up lessons, given throughout a driver's career, aimed specifically at those aspects of behaviour which are likely to have deteriorated.

Perhaps the general point is that many separate strands of thinking in driving research might be integrated by a theory of goal choice. Such a theory allows researchers to relate topics as diverse as skill development, risk-taking and individual differences. As for accident countermeasures, training, selection, propaganda and redesign of the vehicle or traffic environment can all be considered in these terms (Duncan, 1990).

A.1.15 Risk homeostasis theory defended by Wilde (1988)

Main propositions of risk homeostasis theory

Questions as to whether, in which manner and to what extent road users alter their behaviour and their accident risk in response to legal, educational and technological interventions and innovations attract a considerable amount of attention. The answers to these questions have compelling implications for the understanding of accident causation, and thus for the development of rational strategies for the promotion of road safety against various criteria – such as safety per kilometre of mobility, per time unit of exposure to road traffic, or per head of population.

On one hand, the records show that major reductions in the accident rate per kilometre driven have been achieved, in particular road sections or types of roads and in the road network as a whole. The improvements have resulted from interventions and innovations which were not explicitly aimed at reducing the level of traffic-accident risk road users are willing to accept. However, it would appear that these interventions have failed to produce equally favourable effects upon the accident rates per time unit of road-user exposure or per head of population.

On the other hand, there are strong indications that the rates of accidents per time unit of exposure as well as per road user and distance driven can be greatly reduced by interventions which increase the utility differential which people face between having and not having an accident. This may be exemplified by the lasting reduction in accidents per unit distance driven by as much as 86% in a large sample of professional drivers, after they had been promised a bonus for not having a traffic accident (Tschemitscheck, 1978).
In order to explain such contrasting findings, as well as many other observations concerning road-user behaviour, the theory of risk homeostasis has been formulated (Wilde, 1982a,b). This has met with both criticism and support from other researchers.

What the theory of risk homeostasis (RHT) posits concerning traffic accidents may be roughly outlined as follows: at any moment of time, road users monitor the level of accident risk to which they feel exposed and compare this level with the degree of danger they are willing to accept. This preferred or target level of risk depends on their perception of the advantages and disadvantages accruing from their amount and manner of mobility. The actions taken to keep ‘experienced’ risk in balance with ‘preferred’ risk carry an objective likelihood of accident. The actual accident loss incurred by a jurisdiction is the consequence of these actions. This loss, in conjunction with everyday experiences of risk, influences the perceived level of risk among road users who have had no fatal accident. Thus, accident loss and the degree of caution displayed in road-user behaviour are related to one another in a compensatory process, and only those accident countermeasures which are effective in decreasing the preferred level of risk can reduce the accident loss per capita.

The target level of risk is thought of as a construct which people arrive at in an intuitive manner, not as the result of explicit calculation of probabilities of particular outcomes and their positive or negative values.

The average estimation of risk across all members of the road-user population either equals actual accident risk aggregated across behaviours in relation to traffic situation, or deviates from it by a constant (i.e. steady-state) error. This is possibly the most (if not the only truly) speculative (‘axiomatic’) assumption made by the theory, because no information seems to exist which might help verify it, nor any currently available method to provide such information.

The population traffic-accident rate will not drop as a result of regulations mandating the use of seat belts or motorcycle helmets, limits to speed or blood-alcohol levels, vehicle manufacturing standards, the training of road-user skills, or warrants regarding road-side geometry and signalisation, or indeed following any other intervention which does not reduce the target level of traffic accident risk (Wilde, 1982a, 1985a). Interventions of the type described, if effective in changing the frequency of specific behaviours in traffic to which they are addressed, lead road users to display other (for example, non-regulated) risky behaviours in road traffic at a higher frequency. This shift (popularly called ‘conservation of risk’) may take a variety of forms: changes in the frequency of driving, the choice of transportation mode, the amount of road use at night or under
adverse weather conditions, the amount of mental capacity devoted to the driving task, following
distance, driving speed, and so on. However, RHT does not specify in what particular behavioural
form this ‘equifinal’ (Bertalanffy, 1968) compensatory adjustment will take place, but instead that it
will take place and that the accident loss per time unit of road-user exposure or per capita (at least in
the long run) will not be modified beyond the threshold of the just-noticeable difference in accident
loss. The past accident loss of the population and subsequent caution on the part of the surviving
members of the population (i.e. all those who had non-fatal accidents plus those who had no
accident at all) relate to one another in a time-lagged compensatory function (Wilde, 1982a).

Whenever a technical, educational, legislative or whatever intervention – which does not alter the
target level of traffic-accident risk – is introduced, short-term fluctuations in the traffic-accident loss
per capita may occur, but these are eventually eliminated through the closed-loop control process
and the accident rate returns to the pre-intervention level. If road users on average initially
overestimate the intrinsic safety benefit of the intervention, the population accident rate will rise
temporarily. If, at first, they underestimate it, the population accident rate will show a transient
reduction, but in either case, the latter will eventually return to base rate. RHT does not per se
contain an indication as to whether initial estimates err in the direction of overestimation or
underestimation; extraneous information will be necessary for the purpose of predicting the
direction of initial changes (if any) from the pre-intervention accident rate.

Any intervention which does not affect the target level of traffic accident risk and pertains to some
mode or form of road use, but not to others, may cause a shift in accident loss from the addressed to
the non-addressed modes, while the aggregate traffic accident loss per head in the total population
remains at the same level. For example, if travel by car becomes more attractive, a larger
proportion of all accident loss will occur to car occupants; as motorcycle or bicycle riding becomes
less attractive, their proportional contributions to the total traffic-accident loss will diminish.

Discussion of commentaries on risk homeostasis theory

In the ‘zero risk’ model of Näätänen and Summala (1976), driver behaviour is assumed to be
directly related to the level of subjective risk. In most circumstances, subjective risk is perceived
to be effectively equal to zero, so that under normal road conditions traffic participants feel and act
as if they are not running any real risk at all. In other words, there is a threshold for risk
perception, and only if that threshold is exceeded, risk compensation mechanisms are called upon in
an attempt to lower the prevailing risk level (Michon, 1985). The notion of skilful performance
means that many of the cognitive processes occur at a ‘pre-attentive’ level of awareness, require only little information-processing capacity, and are thus not experienced at a high level of consciousness, but they can be called into focal awareness. (Ben-David et al., 1972; Welford, 1976; Anderson, 1982). Similarly, with regard to Fuller’s (1984) conceptualisation of driving behaviour as ‘threat avoidance’, it must be assumed that an above-zero level of perceived threat has ‘adapted out’ from continuous focal awareness through habituation (familiarity). The very act of driving engenders the threat of an accident, thus, the decision to drive implies acceptance of above-zero accident risk.

That Summala may no longer be totally confident of the model he and Näätänen have developed, could be inferred from his recent statement (1986) that pain, like risk, ‘is also sometimes tolerated for important reasons (for example, in the hand of a dentist)’. How is the dental patient different from a driver in a rush while operating a private car, and ambulance or a taxi (Wilde, 1982b)? Similarly, Michon’s (1985) commendation of Fuller’s (1984) concept of the driver as a threat avoider may be questioned for several reasons. First, people do not avoid all threats on the road or anywhere else, although the experience of risk, pain or threat may well be disliked. Second, Michon states that Fuller’s model has a further advantage over RHT, because it is phrased in terms of the behaviouristic paradigm of avoidance learning. It should be noted, though, that Fuller has very wisely refrained from specifying the discriminative stimuli of risk. There are just too many of these stimuli. Given the existence of marked individual differences in sensitivity, any effort to list them would be Sisyphean, and so would any attempt to record the specific behaviour responses in a continuously changing technology (or all the possible pathways of behavioural compensation, unless these are constrained to a small number, as is done artificially under laboratory conditions (Joly and Wilde, 1986; Wilde et al., 1985). Such an undertaking would be as unparsimonious as the never-ending tasks of identifying all discriminative stimuli which create an appetite for food, drink or love, and of making a comprehensive inventory of all specific responses to these stimuli. In contrast, as a typical system theory (Bertalanffy, 1968), RHT makes ample use of aggregation and generalisation and thus attempts to describe the forest, not individual trees: general principles underlying human conduct in the face of accident risk and not primarily highly situation-specific behaviours; its objective is to explain the frequency of occurrence of immediate, ‘material’ (Cownie and Calderwood, 1966) accident causes, not to detail the nature of the latter. Third, what seems to be missing in Fuller’s view of driving behaviour is an ‘approach’ or ‘acceptance of threat’ component. Why should people take to the road at all if their exclusive goal were to avoid the threat of a road accident? For any psychological paradigm to be applicable, it must contain approach as well as avoidance components, hence include a notion of accepted accident risk.
According to RHT, each individual road user tries to keep his/her accident risk per time unit of exposure in equilibrium with his/her momentarily prevailing target level of risk. This represents homeostasis on the individual level. However, no individual can make behavioural adjustments after incurring a fatal accident, but the population of survivors can. RHT proposes that the surviving members of the population do become aware (albeit in a general and quantitatively obscure way) of the magnitude of this loss as well as of other accident costs (personal injury and property damage) by virtue of their everyday experiences which are correlated with these losses: minor accidents, near-accidents and anxious moments experienced on the road by themselves, or by others but communicated to them in personal conversations, and accident reports in the mass media (Wilde and Ackersviller, 1981). These experiences, according to RHT, influence the level of accident risk perceived by individuals who have had no fatal accident and thus their subsequent road-user behaviour.

Various studies have been cited which appear to contain clear evidence that manipulations, deliberately aimed at lowering the target level of traffic accident risk, have led to major reductions in the traffic accident rate per person in groups of road users at whom these measures were addressed (Wilde, 1982a,b; Wilde and Murdoch, 1982; Wilde et al., 1985). In fact, particularly effective in terms of reducing the accident rate per person, as well as per kilometre driven, it seems to have been those safety measures which took the form of authorities extending an incentive conditional upon cautious driving and not having an accident in a specified period of future time. Surprisingly few (Michon, 1984, 1985; Adams 1985; Bonnie, 1985; Kunreuther, 1985) among the many commentators on RHT have reacted to this ‘optimistic’ aspect of RHT: per capita traffic accident reduction can be achieved to the extent that interventions reduce the target level of traffic accident risk in the population.

A few critics seem to doubt the testability of RHT altogether. One of these wonders if it is as difficult to prove as the existence of God (Joubert, 1985). Evans (1986) views RHT as so poorly conceived that all it deserves is to be ignored, and he thus resorts to a deed of censorship which is fortunately somewhat mitigated by its retro-active rather than pro-active nature. In contrast, the great majority of critics acknowledge its amenability to verification by implication of their very act of referring to empirical data which they submit as contradictory evidence.

RHT deals with very large numbers: billions of decisions taken by millions of people over extended periods of time. RHT may be somewhat similar to a theory like the one of evolution (Wilde,
It simply cannot be proven right or wrong in a single experimentum crucis conducted overnight in a laboratory (Wilde et al., 1985), nor even in a single experiment conducted under more realistic, yet necessarily limited, real-life conditions. This is not because the theory if fuzzy or 'axiomatic', but because of practical and ethical complications (Wilde, 1982b).

It would seem fair to deduce from the materials and inferences offered by many commentators that the debate on RHT has been suffering from problems of communication. On the 'stimulus side’ this may be due to insufficient clarity, precision or detailed examples in the way this theory has been presented in the past. With regard to the 'response side’ of the debate, it would seem that much of the criticism directed at RHT has been due to:

a) Careless reading,

b) Lack of adequate consideration for the particular denominator used in the calculations of accident rates,

c) Misinterpretation of the very notion of homeostasis,

d) The erroneous view that the target level of road-accident risk is inexorably fixed and thus cannot be lowered,

e) Misconception that the latter is defined in a circular manner, and

f) That it is due to people seeking risk for the sake of risk,

g) Lack of distinction between pre-attentive and attentive cognitive processes,

h) Failure to distinguish between aversion and avoidance,

i) Lack of appreciation for the typical features of system theory as opposed to atomistic hypotheses,

j) Lack of distinction between intermediate and ultimate criteria of safety,

k) Disregard for macro-economic factors in time-series analyses of accidents,

l) Confusion as to whether RHT is postulated to operate on the individual or collective level,

m) Confusion between cross-sectional and longitudinal deductions derived from the theory,

n) Misinterpretation of the Japanese experience (Koshi, 1985) as at variance with RHT,

o) Oversight of the theoretical and practical relevance of those accident countermeasures which logically follow from RHT, and

p) Various misunderstandings of the due processes of empirical validation of the theory under discussion.

Wilde (1988) makes three final remarks. The first is a quotation from Adams (1985) that risk homeostasis remains a plausible but unproved hypothesis, and the data for testing it remain frustratingly elusive. However, there is an abundance of evidence for the existence of risk
compensation. Only the magnitude and precision of the behavioural adjustment to (non-motivational) safety measures remains in question. Second, field experiments such as the one, by Aschenbrenner et al. (1986), as well as future archival studies monitoring all relevant variables, can provide data which are not so elusive, but are less equivocal and thus more helpful in assessing the validity of RHT and in identifying the limits of its operation, so that this theory may ultimately be deemed acceptable, or more restrictively formulated, or replaced by a better one. Third, accident countermeasures which take the form of reducing the target level of risk have already demonstrated their effectiveness in drastically reducing the accidents rate per person in both road traffic and industry. Further research and development in this direction would seem to be the most promising approach towards this criterion of safety and thus the major challenge to those in the safety research and application community who are more motivated by the desire to add years to people’s lives than kilometres to people’s years.

A.1.16 Reasons for safety measures deemed unsuccessful by risk homeostasis theory

Wilde (1982a) argues that it is possible to reduce accident rate, but that the only method by which this can be achieved is to increase people’s desire to be safe. There is nothing unique about the argument that increasing the desire to be safe may be a significant safety measure. What is unique is the argument that it is the only significant safety measure (McKenna, 1988). The controversy which surrounds risk homeostasis theory is related, not to those measures which are deemed successful, but to those which are deemed unsuccessful. Those measures which are considered unsuccessful include driver education, driver licensing standards, ergonomically designed environments, laws requiring seat belt use and motorcycle helmet use, and more crash-worthy vehicles. In fact, Wilde (1982a) explicitly rules out all interventions except those which affect the person’s desire to be safe. Herein lies the controversy, because most traffic authorities throughout the world have been and are pursuing a range of measures which Wilde argues are unsuccessful. No one is claiming that all conventional measures are successful but controversy surrounds the claim that none of them is successful (McKenna, 1988).

Wilde (1982a) cited a study by Taylor (1964) as evidence in favour of the theory. Taylor had subjects drive on a wide variety of roads, ranging from urban shopping streets to a motorway. While they were driving, changes in their galvanic skin responses were recorded, as were changes in speed. It was found that the galvanic skin response correlated \( r=0.61 \) with the accident rate previously recorded for the roads used in the study. It was also found that the speed at which the drivers travelled was correlated \( r=-0.67 \) with this recorded accident rate. The argument has been
that the galvanic skin response is a measure of subjective risk, which varies directly with the objective risk. In addition, it is argued that drivers adjust their speed in a compensatory fashion with the known accident rate so that, for any road section, drivers will travel slowly through sections with a high accident rate and quickly through areas with a low accident rate.

Should the galvanic skin response be considered a measure of subjective risk? In his review of the electrical activity of the skin, Edelberg (1972) noted that a variety of psychological interpretations have been offered for this response, including ‘defensive arousal’, ‘cognitive activation’, and ‘alertness for information intake’. Several authors (for example, Helander, 1978; Lindholm and Cheatham, 1983) have argued that changes in the galvanic skin response reflect changes in mental workload. How then could a mental workload interpretation account for Taylor’s results? Why should the galvanic skin response or mental workload among drivers correlate with the accident rate observed on the roads they are using? The answer to these questions would appear to be quite straightforward (McKenna, 1988). Road sections imposing the greatest mental workload on drivers will be those where the greatest number of potential conflicts occur; in other words, at road junctions. In their attempt to study mental workload, Brown and Poulton (1961) gave drivers an auditory task which was to be performed either on its own, or while driving in a residential area or in a more busy shopping area. It was found that the auditory task was performed best on its own, less well in the residential area and worst in the shopping area. These results indicate that varying environmental conditions required varying attentional resources. Under this interpretation, the relationship reported by Taylor between drivers’ galvanic skin responses and accident rate on the roads can be explained simply as a reflection of the fact that more accidents occur at junctions where there is a high mental load than on motorways where there is a relatively low mental load (McKenna, 1988).

The two different theoretical interpretations, risk homeostasis and mental workload, focus on very different factors. A mental workload interpretation would focus on the analysis of accidents in order to discover the human information processing errors which produce those accidents. By contrast, a risk homeostasis interpretation would focus attention away from an in depth analysis of accidents towards a more strategic analysis of people’s desire to be safe.

In considering the time period over which risk homeostasis is thought to operate, Wilde (1982a) argues that it may take as long as two years following the introduction of a major safety measure for accidents to return to their preceding level. If accidents have actually decreased in the first year, then it might be argued that this in not evidence against the theory since homeostasis may well
occur during the second. However, since there is nothing intrinsically significant about the period of two years, this parameter can always be varied to protect the theory against disconfirmation. The theory does, in fact, place critical importance on behavioural changes, which are hypothesised to return the accident rate to its preceding level. It is not only justifiable to examine these possible behavioural changes, it is absolutely necessary. Whilst risk homeostasis theorists cite as evidence in favour of the theory the failure of specific safety measures to reduce accidents, this assertion is legitimate only if the mediating behavioural changes can be identified (McKenna, 1988). The reason for this is that safety measures may fail for a wide range of reasons quite unrelated to risk homeostasis theory. In general, if a safety measure is not implemented, or is implemented incorrectly, then it is unlikely to operate effectively. One must distinguish between the ineffectiveness of a safety measure due to a behavioural change following correct implementation and a failure of implementation itself.

In response to a wide range of studies producing results at variance with risk homeostasis theory, Wilde (1986a) has offered a number of methodological criteria which should be met in any empirical evaluation of it. However, the status of these criteria must be questioned. Because they are selectively applied only to studies at variance with the theory, it is difficult to avoid the conclusion that they are offered post hoc and with the sole purpose of dismissing contradictory evidence (McKenna, 1988). The selective use of these methodological criteria is well illustrated by an example in which results from the same study are cited both as evidence for the theory and as evidence against it. It is only when the results are cited as refuting the theory that the study’s methodology is questioned. Thus, Wilde (1981) cited in favour of the theory the Conybeare (1980) study investigating the introduction of seat belt legislation in Australia. However, Wilde (1982b) acknowledged that Conybeare’s results were not consistent with the theory and, at that point, the study was questioned on methodological grounds.

The great strength of RHT, at first glance, would appear to be its simplicity. However, this simplicity is illusory. For example, the role of risk perception in the theory is paramount.

In the risk perception domain, converging evidence from laboratory experiments (Slovic et al., 1977) and field studies (Kunreuther et al., 1978) indicates that people do not have the required sensitivity to low probability events. In addition, it has been found that context effects are important for risk perception (Slovic et al., 1982a). An important factor to emerge from the risk homeostasis field is the complexity of risk perception, which is not simply related to the frequency of accident casualties as risk homeostasis theorists assert. As noted by Slovic et al. (1984), any
theoretical position which relies only on the frequency of accident casualties as its crucial variable will remain unsatisfactory because people’s perceptions are determined by a variety of factors, including perceived controllability and capacity for catastrophic potential. Risk homeostasis theory, with its emphasis on accident casualties alone, is unable to capture the complexity of human risk perception.

Risk homeostasis theory assumes that the single controlling factor is the target level of risk. If a young boy runs across the road to collect his ball and in the process fails to look out for traffic, is the child’s behaviour directed by a particular level of risk? Almost certainly the child’s behaviour is directed simply by the goal of collecting the ball. Accidents frequently occur in similar circumstances, where the individuals concerned are unaware of risk taking (McKenna, 1988). If one considers risk in the framework of decision making it is clear that risk is an outcome of decision-making, but not a goal. The task of driving a vehicle has been considered by Brown (1982b) to consist of six components; route finding, route following, lane tracking, collision avoidance, rule compliance and vehicle monitoring. Risk of an accident will be associated with decisions at all these levels, though this risk may or may not be psychologically represented and it may play only a minor role in directing behaviour. The fact that the vast majority of drivers consider themselves safer and more skilful than the average driver (Svenson, 1981) would suggest that most people think that accidents happen to other people and not themselves. Why, then should their behaviour be directed by considerations of risk? Whilst the risk of an accident may be very much in the mind of the accident researcher, it may not be in the mind of participants in the activity being studied. Most people who engage in activities with some level of associated risk will have successfully and safely carried out these activities on hundreds and perhaps thousands of previous occasions. Under these circumstances, it seems more likely that their behaviour will be directed by task-related events and goals, which have a much higher frequency of occurrence than accidents (McKenna, 1988).

When Shannon (1986) has pointed to the inconsistent claim of risk homeostasis theorists that accident rate is both constant and variable, Wilde (1986b) has defended his position by analogy. Only by switching between a homeostatic theory and a cost/benefit theory can risk homeostasis theorists attempt to account for and predict both constancy and variability of accident rate.

Risk homeostasis theorists argue that the single factor controlling the accident rate is target level of risk. The terms ‘accepted level of risk’ and ‘target level of risk’ are used interchangeable, suggesting that an attempt is being made to use these terms to both describe and explain the
observed accident rate. No clear independent measure of the target level of risk is proposed; thus, it can always be argued that the target level of risk corresponds to the accepted or observed accident rate currently operating. It follows then that, when authors cite work indicating the effectiveness of conventional safety measures (for example, Evans, 1986a; Koshi, 1985; McKenna, 1985a), this evidence can always be countered by arguing that the target level of risk has simply changed. Because of the circularity in the use of the term, this post hoc explanation of contradictory findings can always be offered (McKenna, 1988).

A.1.17 Implications of the zero-risk theory of driver behaviour

How do drivers adapt to risk? Most novice drivers initially feel a sense of uncertainty or fear in many traffic situations, but extensive extinction occurs with increased experience (Näätänen and Summala, 1976). A beginner, when first seated behind the wheel, starts to learn the use of the controls. The complex co-ordination of hands and legs first requires his/her complete attention and conscious control but, gradually, programmes or internal models develop which automatically, without need for conscious control, take care of vehicle handling most of the time. In traffic, a driver further learns how to guide the vehicle along his/her desired path, how to adjust speed, and how to maintain his/her lateral position on the road. At the same time, the driver achieves a feeling of control over the car and surrounding traffic. Perceptual and motor programmes controlling various aspects of driving gradually become more automatised and predictions get better and better. The feeling of uncertainty diminishes as confidence in control skills increases. However, when starting to drive an unfamiliar car, even the experienced driver has to update his/her automatised control models. At first, the driver may either tolerate slight subjective uncertainty, or slow down from normal speed. During this updating phase, conscious control is also called upon more often. Only when internal models have sufficiently adapted to the controls and dynamics of the new car and predictions been sufficiently confirmed by perception, does the driver feel that he/she has regained full control of the task and does the sense of uncertainty disappears.

With increasing experience, the driver learns, just as he/she has learned physical laws, the behaviour of other cars and pedestrians. With experience, the driver thus acquires internal representations of the traffic system and internal models or expectancies in specific situations. As pointed out by Näätänen and Summala (1976), these expectancies are more perception-like and more deterministic than the reality is: they have often largely lost the stochastic aspect of the traffic system.
This is proposed as one basic mechanism which results in: (i) human inability to take into account small stochastic risks within the traffic system, and (ii) extinguishing of the original and very basic fear responses to a variety of traffic situations (Summala, 1988).

If risk is not normally experienced, how then do drivers control risks? Does risk have any important role to play if, as pointed out by McKenna (1982) and Fuller (1984) drivers seldom experience it? Näätänen and Summala (1976) proposed the driver’s safety margin as a critical measure to be controlled. Instead of regulating some risk measure as proposed by Taylor (1964) and Wilde (1982a), drivers are said to be controlling and maintaining safety margins around themselves, as do humans in any potentially hazardous situation. This concept of a safety margin has a deep-rooted evolutionary and individual history. The concept of a safety margin in traffic safety research could be simply defined in terms of the spatial or temporal distance of the agent from the hazard. Quite simple safety margins are exemplified by longitudinal time or space distances in car following; or by lateral separation when passing or meeting a cyclist; or by the distance between the critical and the real lateral acceleration, friction or speed when negotiating curves.

Automation of the driving task and avoidance learning make it possible that most of driving eventually becomes a habitual activity based largely on automatised control of safety margins in partial tasks. No consideration is normally given to risks. However, in most situations, drivers obviously know exactly what they should not do if they want to avoid a certain, or an almost certain, accident. And, whenever necessary, they can focus their conscious attention on estimates as to whether, say, a gap is sufficiently wide to permit merging, or a following distance is sufficient, or the outer lane is free for safe overtaking. Similarly, they can switch to conscious, controlled information processing at any other level of the driving task (Summala, 1988).

Leibowitz (1986) points to the differential impairment of central and peripheral vision at night, which results in directional guidance by peripheral vision being unaffected whilst recognition of hazards in central vision is seriously degraded. Because the ‘primary function of driving’ (i.e. vehicle steering) is working adequately, drivers are not alerted to the need to shift their thresholds for hazard avoidance and, consequently, driving speeds at night are usually too high.

Drivers are quick to use any opportunity provided by new outlets for their motives if the system changes (environmental, vehicular or statutory). The key to effective safety countermeasures is thus to prevent drivers from changing their behaviour in response to system modifications: that is,
to prevent them from satisfying their motives. The psychological implication of the zero-risk theory is that one needs behavioural restrictions in the traffic system: quite contrary to what the proponents of the risk homeostasis theory state (Summala, 1988).

Experience demonstrates that speed limits have immediate safety-promoting effects on actual speed levels. Specifically, they reduce higher speeds and thus reduce speed variance, which further results in a reduction in overtaking and also in close following (Summala, 1980). But what also happened was that these speed limits blocked the outlet for any motivational tendency to increase speeds in response to safety improvements within the traffic system.

Some writers assert that the primary aim of the traffic system is to provide mobility, not safety. This may be true, but there is an optimum level of ‘safe mobility’ and reliability within the transport system which would call for much lower speeds than present limits if driving time, accident and vehicle operating costs were properly calculated (Summala, 1985b). Furthermore, the time costs in these efficiency calculations are accumulated from the time losses of individual drivers for specific trips, typically amounting to only tens of seconds, or a few minutes, in medium-sized European countries. Humanity pays dearly for these seconds or minutes, and for drivers’ freedom to choose their speed, in terms of fatalities, injuries, wasted fuel and extra road-construction costs. In fact, payment is made in terms of casualties for the drivers’ convenience and freedom to travel at their preferred speed (Summala, 1988).

**A.1.18 Fuller’s threat avoidance theory of driving behaviour**

The requirements for a decision to be made about what to do arises whenever alternative actions are possible, even if those alternatives are restricted simply to those of either acting or not acting. Decisions may be accompanied by a conscious assessment of the possible positive and negative outcomes of alternative actions. However, such conscious processing is clearly not necessary for decision-making to take place. Many actions, which appear to require no decision-making once initiated – such as picking up an object, or eating something – may involve complex sequences of decision-making. This becomes readily apparent from the number of conditional statements required for the translation of the behaviour into a computer programme (Fuller, 1988).

One mechanism through which an individual might make preconscious decisions about alternative responses is that person’s conditioning history: behaviours that have been reinforced in the past become more probable under the same stimulus conditions in the future. An example of this is
where overlearned but inappropriate responses occur in an emergency situation. This is not to deny that there may be an illusion of conscious involvement in such decision-making. But, as Aylwin (1985) has expressed it, ‘For much of the time we do things first and reason scuttles along self-importantly afterwards’.

Stimuli in the road environment are rarely intrinsically aversive but become so only as a result of the interaction between what the road user does and properties of the stimuli themselves. Generally it is the driver’s own actions which determine whether or not his/her interaction with the road environment will be punishing. Thus, stimuli in the road environment (such as other vehicles) may be said to have an aversive potential (Fuller, 1984).

Given a discriminative stimulus or warning signal for some impending potential aversive stimulus the driver may either make an anticipatory avoidance response which, if successful, cancels out or neutralises the potential aversive stimulus, or make no avoidance response. The latter includes not only not responding but also making responses which compete with an avoidance response.

Motivation for anticipatory avoidance, competing and delayed avoidance responses

Anticipatory avoidance responses generally involve some form of vehicle control operation. As a result of previous experience such responses may become conditioned to particular discriminative or conditioned stimuli and may be conceptualised as conditioned avoidance responses (CARs). In competition with alternative responses, such responses are reinforced to the extent that they are more frequently followed by rewarding consequences and less frequently followed by aversive or punishing consequences (assuming different rewarding and punishing consequences have equal valence).

A special case where selected speed may compete with anticipatory avoidance responses occurs as a result of experiencing unexpected delays in traffic. Unforeseen delays require drivers to cover subsequent roadway at a faster speed than originally planned and this faster speed may constitute a competing response.

A further situation where an anticipatory avoidance response is not made might occur where a competing or non-avoidance response is intrinsically highly rewarding. A competing response of high speed for instance may itself be rewarding because it increases arousal to a more satisfactory level.
In summary then, the particular pattern of responses followed by a driver depends on the degree of association (or dissociation) between a discriminative stimulus and a potential aversive stimulus and the rewards and punishments for anticipatory avoidance, competing and delayed avoidance responses. The driver’s previous experience is paramount in determining the nature and values of each of these variables.

**Implications for the learner driver**

Learner drivers are more likely than experienced drivers to make delayed avoidance rather than anticipatory avoidance responses. This occurs simply because it requires experience of the road environment to learn the precursors of hazards and to develop the association between discriminative stimuli and potential aversive stimuli. Delayed avoidance responses when brought forward in time become anticipatory avoidance responses.

**A threat-avoidance model of driver behaviour**

It seems likely that where a driver cannot predict a threatening situation with certainty, given a particular discriminative stimulus, and furthermore where a driver cannot predetermine the most appropriate avoidance response to neutralise effectively a potential hazard, he/she may for these reasons make a partial anticipatory avoidance response, rather than a complete response. Avoidance responses are complex, being characterised by at least three major dimensions in which they may vary, separately or in combination (accelerator use, brake use and steering wheel use). Such a partial avoidance response would not be sufficient on its own to avoid a threat should one be realised. However, it would facilitate execution of a delayed avoidance response should such a response turn out to be necessary.

The arousal concept in Fuller’s theory is intended to include possible cortical, somatic and autonomic expressions of activation. In response to expected or actual threat, increases in arousal are within limits adaptive, facilitating information processing and responding. Also within limits, increases in arousal may be intrinsically rewarding (Berlyne, 1969), independent of any rewarding effects mediated through improvements in performance. This implies that drivers may opt for non-avoidance or competing responses in order to boost arousal to a rewarding level (or maintain it if declining). This feature may be particularly characteristic of the more extraverted driver (Eysenck, 1965). Relatively large increases in arousal, however, as well as being aversive rather than
rewarding, may also have the effect of inhibiting cognitive aspects of performance (Broadbent, 1971) and facilitating stereotyped responding, both of which may be maladaptive.

In exploring the relationship between risky behaviour and conditioning history, Fuller (1988) has focused on the conditions under which avoidance responses occur and attempted to summarise relevant results, mainly from animal studies, which have almost completely dominated research in this area. Until there has been an accumulation of human studies the interim conclusion must be that the implications for safety in driver behaviour are not heartening: much of the evidence points to a preference for delayed as opposed to anticipatory avoidance responding, that is, a preference for the more risky behaviour.

If drivers are angry or excited because of situations independent of the driving task, they may misattribute traffic-induced arousal to their general emotional state and fail to make habitual avoidance responses. Drivers may well display feelings of aggression towards other road users by intentionally delaying avoidance responses, thereby constituting a threat to others and requiring them to make compensatory adjustments instead.

In the threat-avoidance model, it is suggested that subjective probabilities refer not to accidents as such, but to the likelihood of some potential aversive stimulus or threat, regarding which some avoidance response may have to be made eventually. It is also suggested that the experience of risk (i.e. feelings such as fear, and anxiety) may arise when the driver becomes aware of a dissociation between his/her actual response and the appropriate anticipatory avoidance response and also when a potential aversive stimulus becomes more imminent (i.e. the more an avoidance response is delayed).

The threat-avoidance model presupposes that drivers opt for zero risk of accident (i.e. it assumes that such events are always aversive) but that they may make anticipatory avoidance, competing or delayed avoidance responses depending on a variety of factors. These include the rewards and punishments associated with each kind of response, the accuracy with which discriminative stimuli are recognised, the subjective probability of a potential aversive stimulus, the effectiveness of avoidance responses when made, and the driver’s arousal condition. The occurrence of unpredictable events will also influence the choice of response. From an objective point of view it is suggested that in the presence of a threat, after a certain point, the more delayed an avoidance response the greater is the risk of accident. However, this concept of ‘risk of accident’ does not
refer to a motivating variable and, as suggested, a host of factors may mediate delays in avoidance responding.

The extinction of learned associations between discriminative stimuli and potential aversive stimuli and in addition, extinction of anticipatory avoidance responses themselves through lack of reinforcement, may account for shifts away from safety in driver behaviour. In principle, these shifts should be observable until punishing consequences ensue to restore both associations and contingencies. Such an oscillating pattern of learning experience may underlie the evidence that is consistent with a homeostatic mechanism in road traffic accidents, as proposed by Wilde (1981), and explain that phenomenon without recourse to the notion that drivers behave in such a way as to maintain some target level of risk of accident which has some value greater than zero.

Fuller emphasises that in accounting for driver motivation the threat-avoidance model does not require drivers to be sensitive to accident probabilities, not to necessarily opt for potentially more aversive than less aversive consequences, nor to have an inevitable dissociation between subjective and objective risk. It is based essentially on a conceptualisation of the driving task as involving learned avoidance responses to potential aversive stimuli and an application of well-established principles of behaviour to the driving situation.

A.1.19 Decision-making, risky behaviour and bounded rationality

Since the authorities responsible for the safety of the traffic system are not able to guarantee a completely fail-safe system, the burden of final control is largely placed on road users themselves. They are expected to keep the system safe. The collective decision makers, such as car designers, road administrators, and policy makers are primarily concerned with the facilitation of road safety by measures designed to minimise risk. Because the goals of the individual road user and the collective decision makers are different, their concepts of risk will differ. Many ineffective safety measures demonstrate a lack of comprehension of the relevant issues in the decision process of the road user. The assumption is that, if road users behaved more rationally, the traffic system would be much safer (Oppe, 1988).

According to Oppe, he feels as rational as an individual road user, almost ignoring danger, as he feels when he is behaving as a professional safety researcher, stressing the point that some of these risks are not acceptable. The question is why does he, as an individual road user, readily accept the risk of a particular journey under conditions of increased risk, such as alcohol consumption, frosted
windows, and bad lighting, yet on the other hand, as a professional, regard these conditions as unacceptable? The dilemma is a fundamental one in traffic-system operation and, indeed, in any system where the collective risk is derived from a large number of individual risks; where the individuals are the actual decision makers (Oppe, 1988).

There seem to be three types of information on the basis of which driver decisions are carried out: (i) present situation, (ii) memory of previous corresponding situations, and (iii) expectations of how the situation will develop (Rumar, 1988). Decisions in traffic often have to be made on the basis of predicted events, rather than in real-time. Since road traffic has a considerable inertia, predictions are useful. Otherwise driver behaviour would be very jerky and speed would be very slow. These predictions may of course, be less or more accurate. They will usually be less accurate the less experience a driver has, both of traffic situations and of the driver’s own capacity to handle those situations. The purpose of information acquisition is to a large extent to check the predictions. Also, predictions are needed in order not to overload the system when traffic conditions are very demanding.

It seems clear that the quality of information employed in risk perception is poor (Brown and Groeger, 1988). The number of potential hazards present in traffic will limit the amount of time which can be devoted to scanning them and evaluating their potential for danger. In addition, speeds and direction of motion in traffic are seldom stable. Therefore, the potential for hazard is continually changing and drivers’ evaluations of hazard must be continually updated, if they are to remain useful in perception of risk. The quality of information on a driver’s ability to cope with traffic hazards will also be poor, since it will be based on memory of past events, which are unlikely to be identical with the current traffic situation. It will also reflect subjective impressions of events, which are unlikely to be consistent over time. Novice drivers fare even worse by this analysis, since they are insufficiently experienced to evaluate hazards adequately and inclined to assess their abilities inaccurately. It seems possible that drivers develop a repertoire of schemata representing the range of spatio-temporal conditions under which traffic hazards and their own behaviour have been experienced, and that these schemata are fitted to the observed traffic scene in order to perceive, influence, or accept risk. This view of risk perception emphasises the importance of the learning process by which schemata are developed, thus identifying the transition from novice to experienced driver as a prime topic for research.

By definition, ‘rational’ road users do not take risks under normal conditions simply because they like doing so (Van der Colk, 1988). They may tolerate risk because it is inseparable from the
achievement of a goal which they are highly motivated to pursue; for example, arrival on time at a destination. However, risk is simply one parameter of the individual’s journey plan, not an end in itself. Risk-taking in traffic may thus be defined as the result of discrepancy between the road user’s self-imposed demands of speed and accuracy and the actual availability of human aptitudes and psychological resources to meet those demands.

Current problems in transport systems resulting from limitations in human information processing appear not to be mainly attributable to perceptual-motor constraints. Human contributions to errors and road accidents seem largely to derive from breakdown in the supervisory and control functions of the road user’s task. Traffic psychologists working in the field of decision theory and problem solving tend to relate such breakdown to the ‘bounded rationality’ of behaviour control (Simon, 1969).

In order to explain this phenomenon, it is necessary to distinguish between the ‘normative’ and the ‘descriptive’ in decision-making (Van der Colk, 1988). The ‘normative’ decision maker has an extensive complete set of relevant information. Furthermore, this exhaustive database is optimally accessed, processed and incorporated into the appropriate decision rules in order to achieve the best possible outcome. ‘Normative’ decision theory thus describes optimal systematic decision-making (Von Neumann and Morgenstern, 1964). By contrast, ‘descriptive’ decision theory deals with the strategies of real human decision makers and the cognitive processes underlying those decisions. It is a psychological theory, the central notion of which is that the bottle-neck of information-processing is localised in the limited capacity of working, or short-term memory; the transmission of information from the short-term to the long-term, and in the information-accessing strategies of long-term memory.

The limited capacity of working memory makes it particularly difficult for people to cope with exhaustive searches of relevant information whilst operating with optimal (and therefore very complex) combinations of decision rules. Consequently, the ‘descriptive’ decision maker has limited (or bounded) rationality, unlike the optimally performing, non-human, normative decision maker. The human decision maker can only continue to function under such conditions of information overload by restricting the input of information and by using simplified models of real-world conditions. This may not be strictly rational, but it is certainly an intelligent method of coping with human limitations. Without a device of this kind, the human decision-making process would produce no output at all under complex operating conditions (Van der Colk, 1988).
In order to further understanding of the principles of bounded rationality, decision-making may be divided into three main stages: pre-decision, decision and post-decision. These may be outlined as follows:

a) Pre-decision defines the problem in terms of alternative choices. Here, bounded rationality could be characterised by, for example, (i) reducing the number of alternative choices by using ‘rules of thumb’, or (ii) using satisficing, rather than utility-maximising, criteria to value the alternative choices.

b) Decision stages involve estimates of outcome probabilities. Here, bounded rationality could be characterised by, for example, biases involving: (i) availability of information, in which the recurrence of more recent events is assigned a higher probability than the recurrence of less recent events, (ii) the ‘law of small numbers’, in which all samples are given undue weight in generalised inferences about the underlying populations; or (iii) a number of other heuristics described by Tversky and Kahneman (1974).

c) Post-decision stages involve ‘learning from experience’. Here, bounded rationality could be characterised by, for example, (i) biases in hypothesis testing about recent events (for instance, accepting confirmatory and rejecting negative evidence for a particular view), or (ii) attributing randomness to complex, but systematic, sequences of events.

Research has shown that even experienced doctors have limitations in diagnostic situations. Often they will consider only a limited set of possible diagnoses and thus test only two or three hypotheses about a patient’s disease. Information processing is then devoted largely to the search for confirmatory evidence supporting the favoured hypothesis. The premature short-listing of hypotheses is often completed on the basis of very few, readily available data, such as the patient’s age, sex and reported symptoms. In making a diagnosis, doctors appear to give a high value to information which confirms hypotheses derived from these initial data and they tend to neglect information which disconfirms their hypothesis. Thus their decision-making takes into account a very narrow range of possibilities and is biased by initial impressions. Computer programmes have been developed to avoid such biases and help the decision maker make unprejudiced medical diagnoses (Van der Colk, 1988).

There appears to be enormous scope for the use of such decision support systems in traffic management, where familiarity with the task can easily introduce a biased hypothesis or short-listing, where the limits of working memory can present real difficulties in information processing, and where the consequences of human error can be equally disastrous as those which may follow erroneous medical diagnoses.
‘Stress’ behind the wheel has become a concern not only of behavioural scientists who study traffic, but of the public as well. Stress, in the sense of mental overload, is seen not only as a subjective and limiting condition about which many drivers complain, but also as a possible cause of disturbances in the traffic system and even of many traffic accidents (Hoyos, 1988). The greatest part of an average driver’s time is probably spent executing routine driving tasks of limited complexity and intensity. However, driving under such conditions makes certain demands on the driver. The individual cognitive and motor tasks which must be performed may not in themselves lead to stress, but the cumulative effect of executing these tasks can lead to overload.

According to Hacker and Richter (1980), psychological strain can be defined as the application of psychological performance capacities when executing delegated work tasks under predetermined external conditions and given individual psychological performance levels. Thus strain is regarded as the process of putting one’s capacities to use in the course of ‘normal’ work activities. Driving a car usually requires only a ‘normal’ application of performance capacities. On the other hand, when talking about ‘stress and driving’ one tends to think of the effect of extremely limiting and threatening influences which a driver needs to cope with. The driver is then confronted with stressors which he/she has to deal with in order to survive.

There is, in fact, sometimes too little and sometimes too much strain for comfort (Hoyos, 1988). Demands vary in intensity. In the case of an overload undesirable strain results, and in the case of an underload, one could speak of a state of deprivation. Undesirable strain, however, need not always be the result of the intensity of the demands made on performance capacities; it is dependent on their duration as well. Undesirable strain is the result of stressors which can sometimes be attributed to a situation (for instance, time pressure or task deviation), but which can also be localised in the person performing the tasks as well (for instance, fears or conflicts).

Most models of stress and strain are based on the assumption that there is a lack of equilibrium between individuals and their environment, which results from a discrepancy between demands and possible ways of dealing with situations. If there is a lack of balance between the activation level and the performance capacity of a driver, or if the demands which traffic makes on performance produce tensions, then there is the risk of an accident, in which the necessity to make decisions under time pressure and using limited mental capacity plays an important role. Strain can,
however, also lead to inadequate emotional reactions such as fear or anger, which require internal regulation.

Under the influence of internalised self-expectations, for example, stress tolerance, an individual looks for a new balance between stress and strain. Coping behaviour is initially the result of cognitive processes and is then controlled by these processes. In general, a person who is in the process of executing a task will verify whether he/she is able to come to grips with effective stressors. If this person finds it possible, for instance to (i) get more information, (ii) use a minimal amount of energy; (iii) understand the situational conditions, or (iv) make predictions and change conditions, then the amount of stress experienced as well as the results thereof will be reduced (Hoyos, 1988).

Determinants and consequences of strain

The driver himself/herself modifies and regulates the effects of stress, which itself leads to a certain amount of strain. The extent to which actions become automatic is an essential personal determinant of the process leading to strain. The more automatic an action is, the more superfluous do various processes of consideration and comparison become. When actions are automatic, demands made on a driver are at a minimum, as well as uncertainty as to whether the situation can be adequately dealt with. One’s overall behaviour while driving a vehicle is under an optimal amount of strain when as many individual actions as possible have become automatic and when many checking and planning processes can take place at higher cognitive levels while these actions are being implemented.

Driving styles can be seen as superposed determinants of driving behaviour. A speed fanatic will try to demonstrate how powerful his/her car is and how well it can hold the road. This driver will experience situations which are different from those which are likely to be encountered by someone who likes easy-going, comfortable driving. This type of behavioural style can sometimes be explained quite well in terms of cognitive style. Participants in traffic situations more or less expose themselves to strain in various quite stable patterns of actions, by choosing situations to which they are willing to expose themselves. Thus, a connection with exposure-based risk exists.

An optimal amount of strain, by definition, also contributes to optimal task execution. Negative consequences are only to be expected when the possibilities for compensation of undesirable strain conditions no longer exist, thus producing a discrepancy between subjective and objective
consequences. Subjective consequences relate to emotional and physiological reactions. Objective consequences are the results of actions; in as far as these influence the driving situation. Inadequate actions which lead to the destabilising of the driver-vehicle system on the vehicular guidance level, or on the stabilisation level, are relevant to safety. Of course, under conditions of undesirable strain, risky decisions can be expected. They precede and result in inadequate actions. Decisions can be risky in those cases (i) where there is not enough time to appraise the situation and necessity for taking action, (ii) when confronted by a complex situation, the driver reverts to simple and perhaps oversimplified strategies, and (iii) where stereotyped and perhaps inadequate patterns of actions become dominant.

Risk exposure

Drivers have the perceptual abilities to identify hazardous objects, hazardous movements of road users, and other hazard-related aspects of the situation displayed. However, there is strong evidence that abilities to perceive hazards improve and change with increasing knowledge and experience (Ganton and Wilde, 1971; Soliday and Allen, 1972). Perceiving and recognising hazards are the bases for safe driving. A driver who sets out to drive a car and thus exposes himself/herself to certain dangers can anticipate the appearance of hazards. Gathering experience in road-traffic situations means learning to anticipate hazards better. But a good ability to recognise hazards does not guarantee safe driving. The consequences of individual hazards, i.e. the possibilities of a collision, must be appraised and individual judgments with respect to traffic must develop into aggregate overall judgments about the hazardousness of a given situation (Hoyos, 1988).

Hazardousness depends on the information-processing load, although many loading situations can be quite unhazardous. However, information load on the road must always have hazard potential, because hazard control requires a certain amount of processing capacity and must compete with other concurrent driving tasks. An important option open to risky behaviour is, without doubt, on the vehicular guidance level (Alexander and Lunenfeld, 1984), the variation of speed, which can clearly be regulated in such a way that the difficulty of the driving task can be compensated for. Clear evidence of compensatory changes in speed can be found. For example, Von Pupka (1977) allowed subjects to engage in other activities while being tested in a driving simulator (lighting a cigarette, eating an apple, using a cassette recorder). They clearly drove more slowly while engaged in these other activities. The more strain the drivers experienced, the more slowly did they drive, and vice versa.
There exist two types of behavioural activities: those necessary for the performance of the task and those which are not directly imposed by the task (Rogè et al., 2001). The latter are called non-specific activities, subsidiary activities (Christol et al., 1979), or collateral activities (Desolvè, 1987). Researchers distinguish between five categories of such behaviours, which can be defined as follows (Cosnier, 1977):

a) ‘Postural adjustments’ are movements of one or several parts of the body in space.
b) ‘Verbal exchanges’ are exchanges which do not include any piece of information about the activity itself.
c) ‘Ludic activities’ are movements implying the manipulation of objects.
d) ‘Self-centred gestures’ are movements of one or both hands towards the body.
e) ‘Non-verbal activities’ are changes which can be observed on the face.

Analyses of these behaviours, which are irrelevant to the primary task, have been carried out mainly in industry. Their number progressively increases with the duration of work, whether it is carried out in daytime (Kishida, 1973), or at night (Delsolve and Preteur, 1986). This phenomenon was also observed in actual working conditions and in laboratory situations, for example, during a monotonous tracking task performed on a computer screen (Phillipps-Bertin et al., 1994). The monotony of the working situation also induces the increase of these behaviours (Kishida, 1973).

In the field of monotonous and prolonged car driving, the variations in the maintenance of position were found to be directly related to low levels of vigilance. When a subject is in a state of low vigilance, the tilt angle of his/her head off the vertical increases (Fakhar et al., 1991).

There is a consensus in the literature about the interpretation of these behavioural variations, as being indications of individuals’ levels of arousal. They cause an increase in the level of activation when a person has to perform a task which is monotonous or propitious to fatigue, because of their distractive effect (Cosnier, 1977). The occurrence of these activities is a defence mechanism which enables the maintenance of performance or of a stable level of output. However, they are not always efficient. There exists a threshold (or and optimal level of vigilance) beyond which performance decreases while the number of non-specific activities continues to increase (Devolvè and Queinnec, 1985).
Rogè et al. (2001) propose that there is a temporal coincidence between the occurrence of the low vigilance phenomena (assessed with alpha and theta physiological indices) and the increase of non-specific behavioural activities. To validate this hypothesis experimentally subjects drove for two hours in the Vigilance Analysis Driving Simulator. The results confirm the interpretation proposed in the literature about the variation in behaviour when an individual performs a monotonous activity. The increase in non-specific activities accurately reflects the variation in the level of arousal.

It seems probable that certain activities allow the driver to reactivate himself/herself. However, contrary to an idea often expressed in the literature, not all behavioural categories have a reactivating effect. Although behaviours such as ‘ludic activities’ could be intuitively associated with the idea of reactivation, the results obtained in the experiment allow Rogè et al. to argue that they do not play this role for the driver. On the other hand, ‘self-centred gestures’ were certainly reactivating, as they occurred frequently when a previous time interval was characterised by an increase in low vigilance signs. Thus, it seems that ‘self-centred gestures’ decrease the number of low vigilance phases during a monotonous and prolonged task. ‘Postural adjustments’ were more frequent during, but also after, an interval with an increase in low vigilance signs. These behaviours may have two functions. During low vigilance periods, the driving task can be more difficult to carry out because the driver has difficulties reacting rapidly and must make extra effort to maintain his/her driving performance. The postural adjustments may constitute manifestations of this discomfort associated with a low state of vigilance. They are probably also carried out in order to reactivate oneself, but they are certainly less effective in this last role than self-centred gestures. The results obtained also indicate that ‘non-verbal activities’ are the only precursory signs of a decrease in vigilance in the context of monotonous car driving.

However, an individual’s self-awareness of which behavioural activities are precursory signs for him/her of an increase in low vigilance, does not allow one to assume that he/she will be able to notice them while engaged in a monotonous task. In the same way, knowing which activities can be reactivating does not allow one to know if the driver will be in a position to voluntarily make use of them to reactivate himself/herself, if necessary (Rogè et al., 2001).
A.1.22 Accident causation and decision-making in emergency situations

Accident causation

Accidents are preceded by long histories containing multitudes of events. These antecedent events can be classified into at least four groups which occur in this order: failure types; psychological precursors; unsafe acts; and breakdown of defences (Wagenaar and Reason, 1990). Reason et al. (1988) have listed a number of general failure types that precede accidents:

a) Hardware defects (wrongly-designed intersections, unsafe car designs);
b) Incompatible goals (speed limits increase safety but incur a loss of time);
c) Poor operating procedures (poor or illogical traffic regulations, for example, on roundabouts);
d) Poor maintenance (roads in poor condition, street lights broken, too many defective cars);
e) Inadequate training (many drivers too young, inadequate driver qualification testing);
f) Conditions promoting violations (unnecessary traffic lights, lack of police control, road repairs causing long delays, insufficient parking space);
g) Lack of organisation (no systematic traffic policy, no systematic collection of accident statistics, no organised reaction to public complaints).

The technical design of the road traffic system allows travellers to trade time against safety. In rail traffic and air traffic, this opportunity does not exist. The general organisation of the road traffic system prolongs this conflict instead of removing it. Any measure which uncouple safety and travel time can be instrumental in removing this important failure type (Wagenaar and Reason, 1990).

Conscious and wilful risk-taking is not in the list. The reason is not that risk-taking may not exist as a cause of accidents, but that risk-taking is a psychological precursor which may result from the general failure types such as incompatible goals or inadequate training. Thus, in terms of safety measures, it will be more efficient to remove the conditions that encourage risk-taking, than it is to keep these conditions and try to convince drivers not to succumb to them (Wagenaar and Reason, 1990).
Decision making in emergency situations

It is often pointed out to researchers interested in emergency situations that they are effectively looking at a phase of an accident on which they can have little influence and that it would be better to study either what happens beforehand, with the aim of preventing the driver from finding himself/herself in such a situation, or to focus on the result, so as to minimise the seriousness of the impact (Malaterre et al., 1988). The briefness of an emergency situation leads some people to false believe that the driver no longer has the possibility of choosing what he/she does, but that he/she simply relies on primary reflexes, which makes all drivers equal when it comes to avoidance.

A fundamental question is why, in certain accident situations where an emergency manoeuvre is possible, does the driver not make the right decision or implement the manoeuvre properly? Over and beyond the interest it may have from a theoretical point of view in understanding perception and decision-making mechanisms under stress and within temporal constraints, answering the question would enable researchers to assess the scope for remedial action against accidents.

For this reason Malaterre et al. (1988) observed actual accidents. The attempted manoeuvres all failed by definition. Most of the time the manoeuvre actually executed by the drivers involved was braking, with or without another manoeuvre. In many cases, a slight sideways movement would have been appropriate in avoiding the accident, but the driver reacted too late or too violently, or tried to combine braking with a sideways avoidance movement, which often resulted in loss of control. In general, drivers’ reactions are too hasty and excessive, which means that either the wheels lock when they brake, or it becomes impossible for them to restore the vehicle to its original trajectory after swerving sideways. This tendency to overreact in emergencies has been recognised for a long time (Davis, 1958).

In a subsequent experiment drivers of varying driving experience were shown slides depicting different conflicts at intersections. The dynamic situation prior to each conflict was portrayed from the front seat of a vehicle and filmed using a video tape recorder. The speed of an intruding obstacle (a van) was modulated to put it on a collision course. The subjects were asked to indicate what action they thought they would have taken and their reason for doing so. The results indicate three conditions necessary for the driver’s primary response to be a sideways movement:

a) Short distance,
b) Certainty about the obstacle’s trajectory,
c) Good visibility.
When they are very close to an obstacle, drivers realise the pointlessness of braking. They start to swerve to one side before reaching the obstacle in the hope that it will stop. A course of action aimed at passing behind the moving obstacle can only be attempted if the driver is some distance away from it. Nearly half the subjects’ manoeuvres were conditional, which is to say that drivers initially braked but, as the situation developed, they had to take a second decision, about whether to continue braking or to veer sideways. Thus, most of the manoeuvre descriptions began with braking. Despite the limitations of this less than realistic experiment, which is not representative of accident situations, it emerges that the excessive use of braking may not be simply a sensory-motor response, but also indicative of a heuristic process (Malaterre et al., 1988).

Perceptual judgments

Malaterre et al. also studied whether drivers are aware that a sideways avoiding action is generally still feasible comparatively close to the obstacle. Subjects did not actually perform the manoeuvre in question, but were asked to press a button to indicate the final limit beyond which it would be too late to either brake or make a sideways movement to avoid an imaginary stationary obstacle.

Estimates for braking and sideways avoidance were significantly different on the part of all subjects. Lateral movement was generally considered feasible for ‘a longer time’, which means that, at a given speed, subjects thought it could be executed nearer the obstacle than braking. Although subjects were experienced drivers, they produced widely differing estimates of feasibility thresholds.

Even if the subjects generally appreciated the gains obtainable from alternative actions, in terms of time and distance, there were considerable variations between them, and their knowledge of actual avoidance possibilities remains vague and irregular. If one adds considerations such as the effects of stress and temporal constraints, one might assume that people do not rely on refined perceptual judgments for taking collision avoidance decisions, but tend to follow simplified strategies, resorting to braking as their initial reaction, except when the obstacle is very close and positioned laterally to their own trajectory (Malaterre et al., 1988).
A.2 Driver Training

A.2.1 Summary

The research field of young drivers is still highly focused on one category of problems at a time, which makes it almost impossible to understand the relations between experience and driver training, experience and motivation, driver skill and risk compensation, or lifestyle and driving style (Gregersen and Bjurulf, 1996). The analysis of driving behaviour factors shows that several dynamic and interactive processes are involved. Driving behaviour and accident involvement are related to different processes such as skill acquisition, self-assessment, information processing, feedback, and motivation. There are varying degrees of knowledge about these factors and their interactions and a great amount of research remains to be done.

There is general agreement in the literature that both age and driving experience correlate with aggregated accident risk for driver populations. The very young and beginning drivers have been classified as groups which are over-represented in crashes, but unfortunately the former is often used as a surrogate for conclusions concerning the latter. Research suggest that licence restrictions or some other form of sanction applied following initial traffic law contraventions could address a substantial proportion of subsequent crash involvement likelihood.

By permitting instructor-supported driving practice from a lower age limit, the intention is to enable young people to gain more experience in driving a car before they acquire a driver’s licence. It is hoped that the extremely high accident risk among young and novice drivers can be reduced. A prerequisite of introducing a lower age limit is that as many as possible are able to take advantage of such a new system. Thus, it is of interest to learn which groups of young people with different lifestyles and socio-economic factors will be making use of the lower age limit. Having a learner’s permit in Sweden is found to be more common among youngsters from white-collar families than among those from blue-collar families. Furthermore, youngsters who are both parent-oriented and friend-oriented practise the most.

Efforts to make novice drivers drive more safely on slippery roads by means of special courses have failed to a large degree. Analyses of the importance given to anticipating versus manoeuvring skills reveal differences between the assessments of instructors and students. Manoeuvring exercises are widely used in advanced driving courses although the main purpose of these courses is to develop anticipating skills. The exercises may give students the impression that manoeuvring
skills are more important than anticipating skills. Manoeuvring exercises also increase their self-confidence and may lead to underestimation of the risks involved, resulting in riskier driving behaviour.

The general problem in the learning process of a driver is that the natural contingencies of the roadway environment are often not adequate to establish and maintain safe driving behaviour. This may be because the contingencies are too difficult to learn, the antecedents of hazards are too unreliable, or behaviour incompatible with safety is too strongly rewarded. One possible solution to this type of problem is to reinforce the following of rules which specify safe behaviours.

A.2.2 Modelling young and novice drivers' behaviour

Extensive research has been carried out to find out why young drivers behave the way they do. The development of theories and driver models have, however, not advanced so far that it is possible to draw reliable conclusions about causes. Without detailed knowledge, it is impossible to make decisions about suitable measures. Much safety work is carried out on the basis of common sense, which is often found inadequate in reducing the high accident involvement among novice drivers. Research in this area has produced or applied theories and empirical findings, pointing out a number of interesting aspects relevant for the behaviour of young drivers. Several studies have shown the importance of driving skills and how young drivers overestimate their capacity (Gregersen and Bjurulf, 1996).

The learning process consists of three main branches. One refers to the initial learning process when becoming a driver and obtaining a driver’s licence. This period is traditionally short and certainly many learners try to minimise it as much as possible to obtain the licence as soon as possible. A typical feature of the first stage of learning to drive is the importance of formal rules and the instructions given by the teacher (Brown et al., 1987). Much of the development of the learning process has focused on this initial period, trying to decide what the best training strategy is. There are several studies on these educational strategies showing that drivers who learn without professional teaching do not have a higher accident risk than drivers who learn from traffic schools (Gregersen, 1994). Another conclusion from research into driver knowledge and skill is that many evaluations of efforts to make the teaching process more efficient have failed to prove any effects on safety (Gregersen and Bjurulf, 1996). It is unknown how the levels of knowledge and skills influence accident risk in itself or whether there are any threshold effects where additional knowledge adds only marginal gains to safety.
The other branches of the learning process are related to the long-term experience where feedback from the traffic is important for the risk evaluation and thus influences the motivational aspects of driving. Experience is also important for the skill acquisition process where behaviour patterns are automated and the mental workload during the novice period is reduced.

A common model for explaining why experience reduces accident risk states that the novice driver passes through several phases according to how he/she can handle the large number of tasks involved in driving. The novice driver faces many new situations, all of them demanding cognitive resources. With time, more and more of these tasks will be automated and the need for free cognitive resources will be reduced (Gregersen and Bjurulf, 1996). Anderson (1982) describes the theory of the acquisition of cognitive skill which includes three stages, the declarative stage, the compilation stage and the procedural stage. According to Anderson, knowledge in a new domain always starts out in declarative form and is used interpretively. Through the compilation stage, the system goes from this interpretive application of declarative knowledge to procedures. By building up procedures to perform specific tasks, a great deal of efficiency is achieved both in terms of time and mental capacity, since the interpretive production requires the declarative information to be represented in the working memory, which may be a heavy burden on the working memory capacity.

The significance of overestimation has also been shown in an experiment to determine how different educational strategies influence overestimation (Gregersen, 1996). The study was carried out to test the hypothesis that skill oriented training would produce more overestimation than training oriented towards the driver’s insight into his/her own limitations. Two randomly distributed groups of learner drivers were given these two types of training and were asked to estimate their ability in a number of driving tasks. Their estimation was compared with their actual ability and thus the reality of their estimation could be calculated. Drivers in the group with skill training were found to have a significantly higher overestimation of their ability, but no difference was found in observed ability. The conclusion from this study was that driver training ought to be complemented with practical training that makes the driver realise his/her own limitation. In another experiment, this strategy for driver training was tested and was found to have a large accident reducing effect compared to the control group (Gregersen et al., 1996). In general, it can be concluded from research that young drivers underestimate risks and overestimate their skills as drivers. Thus, there is a relationship between estimated risk and estimated ability.
When it comes to the relation between motives and driving behaviour, there are two kinds of relevant motives (Gregersen and Bjurulf, 1996). People may want to transport goods from one point to another or may just find satisfaction in driving as such. The influence of these motives on driving is governed by the reinforcement connected with the behaviour. The relation between motives and reinforcement is quite complicated. Most drivers want to drive safely. The problem is to define what this implies in actual behaviour. For the individual driver, it is not so difficult to draw conclusions about driving style, but an individual easily draws false conclusions. A driving style that is statistically dangerous may not be considered dangerous by the driver as an individual. If a driver exceeds the speed limits, the most probable result will be that he/she is not stopped by the police, that no accident will happen and that he/she will arrive at the destination faster. This reinforcement helps the driver to draw conclusions about individual safe driving that counteract safety in a statistical sense.

A.2.3 Learner drivers' practice, socio-economic standing and lifestyle

The importance of a driver's license

Mitterauer (1991) discusses important symbolic events in the transition of youth to adulthood and calls these events a 'caesura of youth', or a form of gradual change in status from the world of youth to the world of adulthood. Obtaining a driver's license is one such caesura for young people, both in their own eyes and in the eyes of others.

Andersson (1987) also discusses the importance of a driver's license for young people's entry into the adult world. Andersson believes that for young people, a driver's license is a major step towards independence and self-reliance from their parents. A driver's license is important not only because it grants permission to drive a car independently and offers a means of identification, but it also serves as a visible symbol of who one is and what one is.

Driving practice, lifestyle and socio-economy

By permitting instructor-supported driving practice from a lower age limit, the intention is to enable young people to gain more experience in driving a car before they acquire a driver's license. It is hoped that the extremely high accident risk among young and novice drivers can be reduced. A prerequisite of introducing a lower age limit is that as many as possible are able to take advantage
of such a new system. Thus, it is of interest to learn which groups of young people with different lifestyles and socio-economic factors are making use of the lower age limit (Berg et al., 1999).

The term ‘lifestyle’ is used in various research contexts to describe people’s attitudes, values, value judgements, opinions and activities. Another frequently cited basic assumption is based on the individual’s need to indicate his/her social affiliation or status. Thus, it emphasises the social and cultural aspects of a person’s life (Gregersen and Berg, 1993). Hermansson (1988) regards lifestyle as referring to people’s actions and a complement to the class structure. It is difficult to place young people in classes since the classes are dependent on working circumstances and positions that have been held in the society.

Hermansson describes the link between the parents’ class position and the child’s lifestyle, in that most middle class children stress family life (parent-oriented), while working class youth emphasise interaction with their friends (friend-oriented). Thus, there is a lifestyle difference based on social class, in that youngsters from blue-collar families to a larger extend represent the friend-oriented lifestyle and those from white-collar families are dominated by parent-orientation. A third category of young people stress both friends and parents. Hermansson views this group as a variant of the parent-oriented lifestyle and as being externally-oriented. Berg et al. (1999) found that the externally-oriented group is evenly distributed over the two socio-economic groups, which is different from the findings of Hermansson, in that the externally-oriented lifestyle was most common in the white-collar families.

Having a learner’s permit is found to be more common among youngsters from white-collar families than among those from blue-collar families. This suggests that the cost of obtaining a learner’s permit is a more unfavourable factor among blue-collar families. Among those who hold a learner’s permit there is no difference found in average amount of hours of driving practice between the blue-collar and white-collar groups (Berg et al., 1999). However, when the amount of practice is separated into lifestyle groups, the externally-oriented group has practised the most. The externally-oriented are alternately being together with parents, friends, and other adults and manage to do this without conflicts. This ability is important for practice with lay instruction since a close co-operation is necessary. In addition, the externally-oriented are more goal and career oriented. Hermansson’s description of the different groups gives a basis from which to explain why the parent-oriented have practised the least compared to the other groups. The parent-oriented group is described as more traditional, moderate and careful. They are short of time because they spend so much time on schoolwork. They share their leisure activities with their parents and thus
do not need a license for these transports. In addition, the relations to the parents leads not seldom to conflicts, which is an obvious disadvantage when practising driving together.

Combining lifestyle and socio-economic status, the externally-oriented from white-collar families have the highest amount of practice compared to the parent-oriented from blue-collar families who practise the least (Berg et al., 1999).

A.2.4 Conflicting goals of advanced (skid) training

In a psychological sense, the basic goal of the skid training is not to promote the use of manoeuvring skills in ordinary driving but to teach anticipating skills in order to prevent risky driving. The special manoeuvring skills taught in the courses are meant to be used only in an emergency (Katila et al., 1996). Because of their increased confidence, drivers do not avoid difficult driving conditions or they can even take on more demanding driving tasks by driving at a higher speed. Results show that skid-training courses have increased drivers’ confidence in their own abilities to drive in slippery road conditions (Keskinen et al., 1992). Gregersen (1996) has reported similar results of young drivers’ increased confidence or ‘overestimation of their own skill’ as a result of skill training in slippery road conditions.

The goal of skid training should be to teach students to use an anticipatory strategy in their driving and to assess their own driving skills in a realistic way. A driver who anticipates risks in traffic can try to avoid them beforehand, for example, by not driving in difficult conditions if it is not absolutely necessary or by driving at a lower speed. An anticipating driver is suspicious of how a situation will develop and is able to look ahead in a longer perspective than a driver whose anticipating skills are inadequate. Drivers should also be able to assess their own skills realistically. Overestimation of or very high confidence in one’s own skills easily leads to situations where the limits of the actual skills are exceeded (Katila et al., 1996).

Why then does skid training lead students to assess manoeuvring skills as so important and increase their confidence in their own skills to an extent beyond their actual skill as Gregersen (1996) has pointed out? One reason for this increase in confidence can be found in the nature of the exercises. Usually they are repeated until students are able to do them correctly and the driving conditions are kept constant. This makes for an easy learning task: the student is learning routines in a constant or static situation. However, the acquired skills can be too limited to be used with success in real traffic where the conditions vary.
It is common belief that good driving skills are for the major part composed of manoeuvring skills (Gregersen, 1996). Unfortunately, manoeuvring skills can also be used to satisfy motives other than safety. Both of these aspects are typical of young drivers, especially young males (Jessor, 1986; Keskinen, 1994; Twisk, 1995).

The feedback given to the students by the instructors is another possible cause for the misunderstanding of the goals of the skid training. As most of the exercises focus on manoeuvring, there is a great temptation to give feedback only about that. For example, evasive manoeuvres on a slippery surface or correct driving in a slippery curve are easily interpreted as means for learning better manoeuvring skills. The original goal of these exercises, that of showing the potential of an anticipating way of driving, has been forgotten, such that lower driving speed makes it possible for the driver to manage a difficult situation. The exercises encourage instructors to give more feedback on manoeuvring techniques in a situation rather than on the students’ abilities to anticipate the situation. Another example of the limitations of the exercises is the brake-avoidance manoeuvre where the driver should try to avoid colliding with an obstacle on the road. The driver has to steer away from the obstacle in the direction (left or right) signalled to him/her with a lamp, flag or hand mark. The learning in this kind of procedure is quite restricted. The exercise is meant to simulate a real brake-avoidance situation, but the only thing the driver learns is to react fast to a sudden ‘stimulus’. However, actual traffic situations do seldom occur unexpectedly. In almost all traffic situations, the driver is able to anticipate how the situation is going to progress. The brake-avoidance exercise is rather limited because it does not favour an anticipatory strategy at all. These kinds of exercises, which have their origin in the tradition of behaviourism, are not well suited to the teaching of anticipating skills.

Learning manoeuvring skills is rewarding in itself. Manoeuvring skills give the young driver a feeling that he/she is capable of controlling the car and thereby get satisfaction from successful operations. The more difficult the operations are, the greater the satisfaction. Rewarding use of manoeuvring skills probably leads to a generalisation of manoeuvring operations from exercises of emergencies to ordinary driving on slippery roads. As a result, the driver may take on more demanding driving tasks, such as driving at a higher speed or accepting more use of sideways skidding.

As a consequence of training or driving experience, drivers’ confidence in their skills increases more rapidly than their actual skills (Gregersen, 1996; Renge, 1995). An experiment conducted by
Gregersen showed that ‘skill’-trained drivers tend to evaluate their skills as better than ‘insight’-trained drivers, after just half an hour’s practice. Positive feedback (easy tasks and repletion), together with the rewarding nature of learning, can increase self-confidence even though the actual skills to manage in real situations have not developed to the same degree.

A.2.5 Crash involvement rates of novice drivers of all ages

It seems evident (Maycock et al., 1991) that both age group and experience are correlated with driver crash risk. Furthermore, there is reason to speculate that the lower levels of competence associated with noviceness may cause the newly licensed to drive less (Laberge-Nadeau et al., 1992). Finally, there is a suggestion that accident risk by age group and experience may be further differentiated by ‘fault’ (Bath, 1992), and there is a potential linkage between the relative frequency of at-fault/not-at-fault involvements and the amount of driving undertaken (Janke, 1991).

All of this underlines the needs to examine in detail the effects of driving experience independent of age group. As immigration to developed countries grows and as more women (for example) and older citizens become licensed at ages beyond the legislated minimum, officials and researchers are increasingly less justified in treating the novice driver ‘problem’ solely as one of assumed youthful inexperience and risk taking. The burden of proof that may be required of legislative or administrative initiatives is clearly illustrated by a decision of the Supreme Court of Canada (1992) in ruling on a complaint regarding ‘rating’ fairness that an insured young driver submitted to the Ontario Human Rights Commission. While the Court did not uphold the complaint since it felt that the insurance industry should be allowed time to develop a better, less discriminatory system of rating, the majority judgment concluded that individual discrimination is not allowable on the basis of statistical averages. These are limitations that face all studies which attempt to link risk with class of driver using only the evidence of aggregate statistics. Within such understood limitations, the purpose of the research reported by Cooper et al. (1995) was to explore the crash-involvement risks associated with novice drivers over a wide age range.

The most interesting finding arising from the analysis of the novice driver data by Cooper et al. is that the higher crash rates associated with novices as compared to general-population drivers were entirely due to higher rates of at-fault, or culpable, involvements. Thus, it would be inappropriate to view novice drivers entirely in a protective or paternalistic sense, that is, to adopt the view that novices are in need of ‘protection’ from others or from the risky environment in which they innocently find themselves. Many novice drivers are the authors of their own misfortune.
Examining in detail the possible contributors to assessments of culpability, researchers found a significant impact associated with the amount and type of driving or travel exposure. Higher culpability proportions are associated not only with those engaging in occasional driving but also with those driving for a living. Lack of experience, as defined by a short licensing time, combined with probable low travel exposure, seem to represent the major problem, and such an indication calls into question novice licensing concepts that propose forced exposure restrictions. However, there are also a group of novices (albeit a much smaller group) whose culpable accident involvements may result from job pressure. They may drive more and they may drive more aggressively than other novices do, because retaining their jobs depends upon meeting time constraints or a daily quota.

The results suggest that licence restrictions or some other form of sanction applied following initial traffic law contraventions could address a substantial proportion of subsequent crash involvement likelihood. If consideration is to be given to designing graduated licensing programmes for all novice drivers which entail initially restricting travel exposure or driving practices, then the results did not indicate any areas for which an explicit safety linkage to noviceness could be made. Furthermore, the higher crash rate for novices than for the experienced argues persuasively for some type of differential licensing treatment. The only question to be answered relates to what form a novice driver licensing system should take. The concept of applying driving restrictions which are gradually relaxed with increasing experience is intuitively appealing, but not necessarily appropriate. Substantive analyses of the specific restrictions to be imposed should support the choice of this path (Cooper et al., 1995).

A.2.6 Difficulties in a driver’s process of learning

Consequence traps

Drivers often become lured into progressively more dangerous speeds on roads because speed is rewarding and they experience minimal punishing consequences. A study of speeds on narrow winding roadways (Svenson, 1978) showed that drivers ultimately learn to travel at such a speed that, should an obstruction occur around the next bend, they would have no chance of avoiding a collision. Such behavioural traps arise because the contingency between a particular rewarding driving behaviour (travelling at a high speed) and a hazardous consequence is improbable and uncertain. Thus, drivers can gamble on the aversive consequence not occurring. Fuller (1991)
calls this kind of trap a 'consequence trap' because the rewarding consequences of the risky behaviour shape and maintain it.

Contingency traps

It is not an uncommon remark from experienced drivers that, when the novice throws away the 'learner' plates after passing the driving test, he/she is only just beginning to learn what safe driving is really all about. Inexperienced divers must learn to discriminate the antecedents of multitudes of hazards from the antecedents of nonhazards (learn to 'read the road') and to learn the contingencies between particular antecedents, particular responses, and their varied consequences (Fuller, 1988b). This learning is difficult due to the complexity of antecedents (which typically include in part the driver's own behaviour), variability (unpredictability) in the contingencies between antecedents, responses, and consequences, and because unsafe behaviour is infrequently punished (consequence traps). A perennial problem for transportation agencies is the difficulty of maintaining safe behaviour among drivers who continuously experience feedback that such behaviour is not necessary (for example, staying within the designated speed limit). Beyond this period of inexperience, despite years of motoring, learning on the roadway never ceases. Drivers are continuously exposed to new and complex sequences of contingencies involving punishment, reinforcement and extinction of both safe and unsafe behaviours (Summala, 1988).

Conditioning traps

Historically it has been the responsibility of road engineers to provide drivers with discrete and unambiguous antecedent stimuli to signal hazards (for example, particular response-consequence contingencies) ahead in the roadway. Such stimuli include warning and control signs, traffic signals, and roadside and pavement markings. However, albeit in some countries more than others, warning signs are sometimes inconsistently located, identifying, for example, a dangerous bend in the road at one point but being absent at an equally dangerous bend further along the same road. This problem is even more evident in relation to the placing of 'road work ahead' warnings which are too often left in place for days, weeks, of even months after completion of the work. It is thus hardly surprising that only about half of the drivers approaching a construction and maintenance zone report reducing speed on seeing signs instructing them to slow down (Gardner and Rockwell, 1983). Over 20% said they waited until they could actually see the hazardous construction work itself. Sometimes the same warning sign signals different contingencies, such as an easily negotiated bend at one point, but a definite hair-raiser at another. Sometimes warning signs
identify contingencies which are only very rarely appropriate, such as the ‘school’ sign for which
the warning is relevant only when children are in transit to and from the school. It is to be expected
that drivers’ behaviour is inadequately controlled by such antecedent stimuli.

What is happening in these instances is that antecedent stimuli introduced to enhance safety are
failing in their function because they often do not effectively discriminate safe from unsafe
situations. As a consequence, appropriate hazard avoidance responses do not come under their
control, responses which in a safer world would become conditional on particular signs appearing in
the road ahead (Fuller, 1991).

The distinction between consequence, contingency, and conditioning traps is somewhat artificial
because all arise out of elements of the same process, the process of learning. Their common
message is that road accidents do not just happen: drivers learn to have them. The point of
distinguishing between them is to highlight the relative importance, under various conditions, of
particular features of the learning process. As described by Fuller, these features include the
effects of rewards on unsafe driving behaviour, the difficulty of learning contingencies in the road
and traffic environment, and the failure of discriminative stimuli to bring safe driver behaviour
under their control.

The general problem is that the natural contingencies of the roadway environment are often not
adequate to establish and maintain safe driving behaviour. This may be because the contingencies
are too difficult to learn, the antecedents of hazards are too unreliable, or behaviour incompatible
with safety is too strongly rewarded. As one possible solution to this type of problem, Skinner
(1988) has argued that one needs to reinforce rule following: in this instance the following of rules
which specify safe behaviours. This strategy has had demonstrable success with safety belt
promotion, where natural contingencies are not very effective at maintaining the desired behaviour
(Geller et al., 1982).
A.3 Drivers’ Perception of Speed

A.3.1 Summary

Research conducted to ascertain the effects of choice of speed by drivers cover a wide scope ranging from the effect of driving speed on reaction time, experiments to establish the mechanism by which drivers estimate their own velocity, and the ability of drivers to judge the velocity or speed of oncoming vehicles for passing manoeuvres. Other issues concern the effect of new technologies for speed control on driving behaviour and the important question regarding car following and choice of headway.

A.3.2 Effect of driving speed on reaction time

For many years the effects of driving speed on traffic safety have been debated, with an unresolved assumption that drivers will find it more difficult to stay alert at slower speeds, Lisper and co-workers developed a subsidiary auditory reaction time task to be presented during car driving on a motorway (Lisper et al., 1971). The validity of this task as an index of driver performance was studied in a series of investigations. In one such study, Laurell and Lisper (1978) investigated the correlation between performance of this task and detection distance to an obstacle as a function of time on task during a three-hour driving session at night. The correlation between the group averages of these two measures was significant \((r = -0.78)\). The average within-subjects correlation was -0.47, also a significant value. These results were interpreted by the authors as supporting the validity of the reaction time task as an index of driver performance.

Tőrnros (1995) used a very similar subsidiary reaction time task. The speed factor consisted of 70 km/h, 90 km/h and 110 km/h. Subsidiary reaction time during car driving was significantly slower at 70 km/h than at 110 km/h. The result for the short reaction-time test presented before and after the car driving session showed no after-effect of the speed factor. There was only a very weak tendency for degradation of performance at the slowest, compared with the fastest speed.

Subjects rated themselves as having felt more energetic when driving at the fastest speed compared with the slowest speed. For the arousal factor the tendency was similar, but no significant effects were demonstrated. For the stress factor no clear tendency in any direction appeared. The rankings after the last test sessions indicate that the subjects on the average had felt sleepier or more
tired when driving at 70 km/h compared with 110 km/h. Therefore, it is evident that the effects of the speed factor appeared only during car driving, with no after-effects being demonstrated.

In a discussion of possible mechanisms behind the demonstrated effects, it should first be pointed out that the driving times were very different for the different speed conditions, an inevitable consequence of the choice to keep driving distance constant (200 km). The interesting issue, that of separating the effects of speed and driving time, could not be resolved in the study. An attempt was, however, made to extract some data bearing upon this issue. To this end, reaction time was studied as a function of driving time, instead of driving distance. No effect from speed appeared in this case where the reaction times from 2 hours of motorway driving in each speed condition were analysed.

Differences in the driving situation may also have had some effect on the outcome of the study. It may be assumed that it was least monotonous at the fastest speed, illustrated by the fact that the subjects on the average made 35 overtakings at the fastest speed, compared to an average of 0.5 overtakings at the slowest speed.

What relevance the findings may have in a real traffic situation can be discussed. The first thing to take into consideration is the size of the effect. The difference between reaction time at the fastest speed, compared with the slowest speed, was 15 ms on the average. This difference should be compared with what has been found in other studies. Lisper, Laurell and Van Loon (1974) found that the reaction time on a similar task had increased by approximately 200 ms just before the subjects fell asleep at the wheel. It is further necessary to relate the findings to the large differences between driving speeds regarding reaction distances and braking distances. The reaction distance increases linearly as a function of speed, whereas the braking distance increases more or less quadratically as a function of speed. The braking distance will therefore be nearly 2.5 times as large at 110 km/h compared with 70 km/h. A similar relation exists between speed and motion energy. Whatever the consequences of these conditions might be with regard to traffic safety, the findings indicate that the possibility of counteracting these consequences by being somewhat more alert at higher speeds are not great. With the small degradation of reaction time found during 200 km of motorway driving it is improbable that reaction distance would be affected to any great extent but would still be larger at the highest speed.

In the study by Törnros, effects of driving speed were examined when driving on a motorway having the speed limit of 110 km/h in all three speed conditions. If a similar study were performed
on motorways having different speed limits, it is probable that the interaction with other road users would be more similar in the different conditions than was the case in the present study, and hence any possible stimulating effect of the differential between own and others’ speed would be minimised.

A.3.3 Estimation of vehicular velocity under time limitation and restricted condition of observation

Salvatore (1967) reported on an investigation of the ability of subjects to estimate the velocity of the vehicle in which they are travelling. In real world, the rapidity with which an estimate is obtained is of importance. Especially in those emergent traffic conditions which call for a reappraisal, the estimate which takes longer is worth less if the two estimates are of a given accuracy and reliability.

The question of whether the velocity of moving objects is perceived directly or by a cognitive operation relating perceived spatial displacement to perceived duration is pertinent. In the periphery, angular velocity is high and movement obvious so that movement of the vehicle through the environment can be apprehended directly, whereas in the fovea movement is slower so that velocity may have to be computed from an estimation of distance traversed and a separate estimate of elapsed time. The latter is, of course, more time-consuming.

The experiment strongly indicated that peripheral visual stimulation is more conducive to accurate speed estimation than stimulation of the frontal field. The immediate explanation for this result appears to be that angular velocity is much greater in the peripheral than in the frontal field. For example, under conditions of the experiment, at a vehicular velocity of 96 km/h and 25° of field available, the maximum angular velocity is 81° per second frontally and 1080° per second peripherally. Angular velocity in both fields is directly proportional to vehicular velocity but since it is greater in the periphery by more than one order of magnitude, it is more accessible to scaling. In addition, the perception of motion in the frontal field is detrimental to steering or tracking performance, the possibility exists that in normal driving the attention tends to focus on the fovea, that aspect of the frontal field which has the smallest magnitude of angular velocity.

Wohl (1961) indicated that the proper steering input is a complex function of vehicle dynamics and is inversely proportional to the speed or square of the speed of the vehicle depending on the primacy of the visual cue used as reference on the roadway. Poulton, in analysing tracking as an analogue of the driving task, emphasised that speed anticipation, the decision aspect based on the perception of or inference about the stimulus movement, is of importance in steering. Similarly,
Cumming (1963) stated that the ability to programme ahead with speed control is the mark of a developed skill.

Thus for the single-car situation, the appreciation of velocity plays a role in the prediction of future states of the system which involves co-ordination or response programming and the elimination of reaction time limitations. Therefore, steering control is intimately connected to appreciation of velocity. Additionally, multiple car manoeuvres such as car following and passing always involve the projection of one’s vehicle into the future.

A.3.4 Scaling of relative velocity between vehicles

Mathematical models of car following assume that the following driver perceives veridically the relative velocity between his/her own vehicle and that ahead and makes decisions based on this relative velocity. In the same way, when making overtaking decisions, a driver must be able to estimate the speed of the oncoming vehicle in order to make a valid decision. Various authors have shown that the perceptual information required by the driver (Hoffmann, 1968; Hoffmann and Mortimer, 1994) is in the visual angle subtended by the lead vehicle at the eye of the following driver and the rate of change of this visual angle.

For valid decisions to be made during overtaking or car following, it is first of all necessary that the approach of the vehicle is perceptible, not just from the change of distance over time (or the associated change in the subtended visual angle), but directly through the perception of its relative speed (through the rate of change of subtended visual angle). This requires that the angular velocity subtended by the observed vehicle be above a threshold value of about 0.003 rad/s (Hoffmann and Mortimer, 1994). When the rate of change of the subtended angle of a lead vehicle exceeds the threshold value, humans have the information available to subjectively scale the relative motion between two vehicles.

Possibly the earliest work which studied the ability of drivers to make judgments of relative speed between cars was that of Olson et al. (1961). They found that subjects were able to do little more than detect whether the headway was increasing or decreasing. Accuracy of judgments increased as headway decreased and when the gap was closing. In these experiments, the maximum subtended angular velocity was about 0.001 rad/s, well below the average threshold value.
Other work has concentrated on cases where the observer was stationary and the vehicle was approaching, or the observer was in a moving vehicle and has to estimate the speed of the vehicle (Häkkinen, 1963). In nearly all cases over the experimental range of 30-90 km/h, the estimated speed was less than the actual speed. The underestimation increased with distance of the vehicle at the time of estimation, to the extent that estimates were likely to be independent of actual speeds and quite unreliable. Analysis of Häkkinen’s data (Hoffman and Mortimer, 1996) showed that at the largest distance (300 m) the angular velocity subtended by the approaching vehicle was less than the threshold value. This may account for the poor estimates at this distance. Denton (1963) reported similar work with an observer moving with the vehicle and found that subjects were able to scale the speed of the vehicle, but that this scale was speed dependent.

When both vehicles are in motion, there is a difference in the available visual information from that with a stationary observer, particularly as there is none available from the motion of the second vehicle relative to the environment. Farber and Silver (1967b) reported a series of experiments for judgments in an overtaking situation and found that drivers can make reasonable estimates of the distance of an approaching car but cannot judge and take into account the speed of the oncoming vehicle when making decisions. Recalculation of the data of Farber and Silver (Hoffman and Mortimer, 1996) in terms of actual and apparent angular velocities indicated an angular velocity threshold of about 0.003 rad/s, consistent with other experiments. When this threshold was exceeded, there was some relationship between perceived and actual speeds. Under most conditions where an overtaking decision is made, the researchers found that the threshold passing distance adopted by drivers tends to remain constant regardless of oncoming car speed. Drivers were not able to obtain information about the oncoming car speed due to the subtended angular velocity of the approaching car being below the threshold value.

Data of Björkman (1963) for an overtaking situation showed similar results, which are also in agreement with data reported in Rumar and Berggrund (1973) who stated that drivers cannot estimate the speed of the oncoming car. Being in that vague position, they seem to assume that the speed of the oncoming car is the same as their own speed and consequently base their decision mainly on the estimated distance to the oncoming car. This is an obvious outcome of the angular velocity of the approaching vehicle being below the threshold value – no speed information is available to the driver of the overtaking vehicle and thus the best estimate that can be made is that the speeds are the same.
It is apparent from the above that information about threshold angular velocities and scaling of relative speed is irrelevant to the problem of overtaking, as this information is not visually available to the driver when the overtaking decision is made. However, with an assumed threshold angular velocity of 0.003 rad/s, the headways and approach speeds are in the range commonly experienced during car following in a traffic flow (Hoffman and Mortimer, 1996).

These results have obvious implications for traffic flow models where human factors are to be included. For example, drivers are unable to respond to the change in headway unless they are able to perceive that a change of headway has taken place. Perception of this change will occur either (i) through the vehicle spacing changing to such an extent that the just noticeable difference is exceeded (Mortimer, 1990), or (ii) by detection of the subtended angular velocity. Depending on the headway between the vehicles, the model should include one or other of these dead zones in which the visual system cannot detect that spacing changes are occurring.

In most situations of car following the threshold for noticing the change in headway will be exceeded before that for relative velocity. Mortimer (1990) has noted that the time for collision will be critically low, in most instances, if the headway is decreasing when the angular velocity threshold is exceeded, and then drivers must respond immediately to avoid a crash.

All of the above experiments have been performed by simulation in the laboratory, without full visual stimulation to the subject, such as might be obtained from peripheral vision. Other research indicates that the use of such a simulation may not be a problem, since it is largely the visual information related to the changes in subtended angle that is used for such estimations (Hoffmann and Mortimer, 1994). Peripheral vision may be useful for estimating absolute speed of a vehicle, but when both vehicles are in motion, there is little useful visual information to be gained from the relative motion of the vehicles, which is largely in the central visual field of the following driver.

A.3.5 Knowledge of oncoming car speed as determiner of driver's passing behaviour

The purpose of a study by Farber and Silver (1967b) was to examine the effect of increased information about oncoming car speed on driver judgment in accelerative passes. An accelerative pass is one in which the overtaking driver starts the pass from a close following position with little or no speed advantage and must accelerate to complete the manoeuvre. In such a pass, sight distance, legal passing zone boundaries or oncoming traffic limits the passing opportunity. Where an oncoming car (OC) is the limiting factor, the passing driver must take into account his/her own
speed and the speed of and distance to the OC to make a valid passing decision – that is, to pass when it is safe to do so, and not to commence a pass when it is not safe.

It appears that since a driver has first-hand phenomenal and metric knowledge of his/her own speed, and is able to discriminate distance with reasonable accuracy, much of the variability in passing judgment is associated with poor judgment of OC speed.

Their series of experiments produced a number of clear-cut results. It may be concluded that drivers are able to make good judgments of the distance to an oncoming car; that at normal passing distances drivers do not respond appreciably to oncoming car speed; and that drivers are able to make good use of verbal knowledge of oncoming car speed in making passing judgments.

It would appear then that one way to improve passing performance is to provide some information about oncoming car speed. The data showed that a marked reduction in passing time gap variance and hence, safety margin variance can be achieved. The mean passing time gap and resultant mean safety margin adopted by a group of drivers may not be affected by the provision of oncoming car speed information. Nevertheless, a reduction in variability in passing gap acceptance and safety margin can have important consequences for both safety and throughput. Lower variance means that more drivers will pass when they should and fewer will pass when they should not.

A.3.6 Choice of time-headway in car following and the role of time-to-collision information in braking

Close car following has been associated with traffic accident involvement. These accidents are usually attributed to maintaining insufficiently long headways and/or to inattentive driving resulting in responding too late to a deceleration of a vehicle in front. For investigating this phenomenon, Van Wisum and Heino (1996) defined DHW as the bumper-to-bumper distance between the lead vehicle and the following vehicle. THW is the time interval between two vehicles in car following, calculated as DHW divided by the speed (in m/s) of the following vehicle.

Several factors have been identified that influence the choice of THW. Choice of THW has been associated with personality factors by some authors. Sensation seeking as a personality trait is assumed to be related to risky behaviour (Zuckerman, 1979). Heino et al. (1992), using a realistic car-following task, reported a smaller preferred THW (THW_{pref}) for sensation seekers than for sensation avoiders. Other authors have stressed the importance of task-related factors with regard
Fuller (1981) studied THW of truck drivers in convoy situations. During the late shift, covering a large period of driving in the dark, THW_{pref} was significantly larger than during daytime driving. This was explained as an effect of visual conditions. Brookhuis et al. (1991) reported an increase in THW when using a car telephone while driving, which can be regarded as an additional task competing for attention. This suggests that the driver is aware of effects of task demands on the ability to detect a deceleration of a lead vehicle and adapts THW accordingly.

Ota (1994) studied THW while drivers were required to drive with a speed of 50, 60 or 80 km/h and follow under different instructions such as ‘follow at a comfortable distance’ and ‘follow at a minimum safe distance’. No effects of speed on THW were found while instruction significantly affected choice of THW.

An important skill that has been associated with the initiation of braking relates to the perception of time-to-collision (TTC). TTC is defined as the time required for two vehicles to collide if they continue at their present speed and on the same path (Van der Horst, 1990). TTC is computed as \( \frac{D_{H/W}}{V_r} \), where \( V_r \) is the relative velocity or speed difference, which must be larger than zero.

Hoffmann and Mortimer (1994) found that both estimated TTC and standard deviation of estimated TTC were linearly related to actual TTC. They reported an underestimation of TTC of 20% on average, while other studies typically reported an underestimation of around 40%. This better performance in TTC estimation was attributed by Hoffmann and Mortimer to the fact that in their experiment both vehicles were in motion, while other experiments typically measured estimated TTC to a static object. The studies on TTC estimation give substantive evidence for underestimation of TTC and for individual differences in the ability to accurately estimate TTC.

The hypothesis that THW_{pref} is consistent within the driver and the hypothesis of constancy of THW_{pref} over speed during steady-state car following were confirmed for the range of speeds examined by Hoffmann and Mortimer. The brake reaction of drivers was analysed in order to investigate whether differences in THW_{pref} during steady-state car following are related to differences in braking performance and underlying skills. Since THW during steady-state following represents the time available to the driver to give an appropriate braking response in case the lead vehicle decelerates, THW may be the result of an adaptation of the driver to individual differences in braking competence. Braking performance was assumed to be related to the ability to perceive time-to-collision (TTC) and the ability to generate an efficient braking response, depending on the criticality of the situation. The initiation of braking, as measured by brake-
reaction time (BRT) was strongly related to TTC at the moment the lead vehicle started to brake (TTC\(_{10}\)) and thus to criticality. This strong relation was apparent between subjects as well as within subjects. This conforms to the suggestion in the literature that TTC information is used by the driver to judge the moment to start braking. However, drivers with a smaller THW\(_{\text{pref}}\) during steady-state following start to brake at a lower TTC, i.e. when the criticality is higher. This suggests a different TTC criterion for the initiation of braking, depending on preferred time-headway. Although the initiation of braking was very sensitive to TTC information, there were no differences between short followers and long followers in sensitivity of BRT to TTC\(_{10}\). Thus, the hypothesis that differences in THW\(_{\text{pref}}\) during steady-state following are related to the ability to accurately perceive TTC was not confirmed.

The minimum TTC during braking was smaller for short followers. This indicates that a collision was more imminent for short followers than for long followers. There were however differences in the control of braking. First, short followers pressed the brake pedal to a higher maximum, resulting in a larger deceleration. Second, for short followers the intensity of the braking response was more strongly dependent on the criticality at the moment the lead vehicle started to brake. Short followers are better able to programme this response to the appropriate level, depending on criticality. At the moment the lead vehicle starts to decelerate, the driver does not know how strong it will decelerate and for how long. Therefore, visual feedback during the braking manoeuvre is important for continuously adapting the braking response to the required level. The programmed braking intensity may then have to be adjusted to another level depending on the development of criticality in time. The moment of maximum deceleration (t\(_{\text{DECmax}}\)) was assumed to indicate when the driver knows a collision will be avoided. A closer correspondence in time with the moment of minimal TTC (t\(_{\text{TTCmin}}\)) suggests a better ability to adjust the control of braking to requirements. In this respect, the third difference was found between short and long followers. For short followers the absolute difference between t\(_{\text{DECmax}}\) and t\(_{\text{TTCmin}}\) was smaller, indicating a more efficient braking control where the timing and intensity of braking is better tuned to the development of criticality in time during the braking process. An alternative explanation may be that short followers had to generate more efficient braking responses that were better tuned to criticality because criticality was higher for them to begin with. In other words, they may have been forced to perform better.
Modelling car-following behaviour

To explain car-following behaviour a range ($R$) versus range-rate ($Rdot$) phase space is a useful construct. For example, the driver is concerned with the range being too small as well as with rapid closure as indicated by a negative range-rate. A model by Fancher et al. (2001) uses perceptually determined boundaries approximately as indicated in the following range versus range-rate diagram.

![Figure A.1 Range ($R$) versus range-rate ($Rdot$) phase space (Fancher et al., 2001)](image)

These boundary lines define six regions as depicted in Figure A.1. The horizontal line near the centre of the diagram represents the desired comfortable range for the speed of travel. Region A is considered the most stressful region such that the driver may be willing to brake hard if the situation is characterised by range and range-rate co-ordinates in this region. In region F the driver is not too close but the range-rate is of concern. The driver is expected to commence slowing in this region. Region B is a transition region between $Rdot<0$ (closing) and $Rdot>0$ (separating). In this region the driver has difficulty perceiving range rate, but knows that the range is too short. Examination of test data indicates that drivers tend to brake in this region but they usually get off the brake as soon as they detect that $Rdot>0$. Proceeding counter-clockwise, consider how the driver chooses to change speed in region C. The range is too close in that region, but the vehicles are separating. The driver knows that the range will increase if speed is simply maintained or reduced slightly. In region D the driver is not too close and is able to perceive the rate of separation of the lead vehicle. In this region, it is reasonable for the driver to try to catch the lead vehicle if it is not going too fast. It is conceptually problematical to develop a simple qualitative rationale for describing driver behaviour in region E. The driver is not threatened by short range or by closing too fast. In this region, other factors may be more important to the driver than closeness or closing threats. For example, the reason for keeping range small might be to keep another vehicle from cutting-in.
There is need for a seventh region in which $R$ is so large that the driver is simply unconcerned about the preceding vehicle. In this region, the driver modulates the throttle to get the desired speed of travel.

### A.3.8 Close-following drivers on two-lane highways

Higher accident and violation rates per driver among close-followers may be due to higher mileage. A possible causal link simply goes through higher target (desired) speed level (Summala, 1996, 1997). The variation in desired speeds necessarily results in the need to overtake slower vehicles. Drivers with a higher target speed encounter slower vehicles and have to overtake more often. On two-lane roads particularly, close following may largely result from this desire to overtake and waiting for an opportunity to do so. Summala (1980) found that a long overtaking prohibition on a two-lane highway reduces short headways, and Summala and Vierimaa (1980) showed that the shorter the following distance, the more left the average lateral position is with respect to the car ahead (for left-hand driving). On multi-lane freeways, too, fast drivers often force slower ones to leave the faster lane by driving very close behind them.

There are two kinds of close following, one temporary and situation-specific and another more consistent steady-state following, and Rajalin et al. (1997) hypothesised that whenever a faster driver reaches a slower vehicle, he/she chooses the headway differently depending on whether he/she plans to overtake the car or not. Frequent overtaking may lead to habitual close following in other circumstances as well.

Active overtakers – people who drive more and faster than the average and follow other vehicles at shorter distance when waiting for an opportunity to overtake – may be more alert, and may also have more practice in managing with slower vehicles (Bjornskau, 1996) to the degree that their motor control to a decelerating lead vehicle is faster (Van Winsum and Heino, 1996; Van Winsum and Brouwer, 1997). ‘Safe’ far-followers may instead be less alert and may also share time and attention between driving and in-car tasks which results in problems in detecting lead car braking; as a matter of fact, the detection of a lead car’s braking even gets more difficult with increasing distance (Lamble et al., 1996) Other things being equal, including attention, close following very likely increases accident risk, while it also get drivers in front anxious and irritated (Hämäläinen, 1988).
Various attempts to reduce speed, such as traffic calming, enforcement and publicity campaigns have had some degree of success, but are often localised in their impact (Casey and Lund, 1992; Corbett, 1995; Teed, Lund and Knoblauch, 1993). A global speed reducing measure, Intelligent Speed Adaptation (ISA) is being developed and is undergoing field-testing in the UK, the Netherlands and Sweden. Fundamentally, ISA attempts to reduce drivers' propensity to exceed the speed limit.

ISA systems can operate in a number of ways. Firstly, they can provide advice to drivers regarding the speed limit for the piece of road on which they are travelling. Additionally, drivers may have the option to engage a voluntary ISA that prevents them from exceeding this speed limit. There are also mandatory versions of ISA, which do not allow the driver to disengage the system. Finally, a dynamic ISA system could limit the maximum speed of a vehicle using real time information concerning the appropriate speed limit for the road and weather conditions.

Using a driving simulator, variants of an ISA system were evaluated in terms of driver behaviour, with particular focus on any safety critical changes that arose (Comte, 2000). A number of changes in behaviour were found across a broad range of behavioural measures. The speed choice data suggests that the ISA systems have little impact on mean speeds. However, the effects of the ISA systems were more prominent at specific locations, such as village entry. It is at these locations in real life, where drivers have difficulty in adapting their speed to the recommended limit.

As in previous work (Persson et al., 1993; Comte, 1996), it was found that when using ISA, drivers' gap acceptance behaviour altered. The mean accepted gaps reduced in size, by a maximum of 1 sec. This suggests that drivers were exhibiting riskier behaviour, however, it is not clear whether this shift in behaviour would contribute to increased accident likelihood, as the threshold for safe behaviour is difficult to define. Thus, this was not viewed as a negative outcome by Comte (2000), but should merely alert researchers to the possibility that behavioural adaptation can occur.

Drivers using either the Mandatory or Variable ISA systems were becoming increasingly frustrated on increased exposure to the system. This frustration led to drivers becoming impatient and wanting to gain lost time. This impatience was then exhibited in the smaller gaps taken. In fact, total journey time did not increase under ISA, although drivers may have perceived it to have done so.
There were also observed changes in car-following behaviour. Close following (less than 1 sec) increased in both urban and rural areas. When driving behind a slow moving vehicle (with no opportunity to overtake), it appears that drivers using an ISA system were more likely to want to maximise their speed and thus engaged in close following. Interestingly, in contrast to the gap acceptance results, it appears that drivers with ISA were not so impatient that they needed to engage in more overtaking manoeuvres.

With regards to driver awareness and the ability to react to changing road conditions, the findings were mixed. On the positive side, no evidence of loss of vigilance was found using a choice reaction task and a potential collision event. This does not support the anecdotal criticism that ISA might induce drivers to ‘switch off’. Some evidence has been found of these phenomena for Adaptive Cruise Control systems (Richardson et al., 1996) and also in previous ISA experiments (Comte, 1998). In this latter study, it was found that drivers who used a Mandatory ISA system were less likely to adapt their speed to prevailing weather conditions. When visibility was reduced, these drivers continued to drive at inappropriate speeds, and more importantly at speeds higher than even those in a baseline condition.
A.4 Lane keeping for Straight and Curve Driving

A.4.1 Summary

In order to understand and model both straight lane keeping and curve driving, the construct of time-to-line-crossing (TLC) proved to be very useful. TLC represents the time a driver has available to neglect path errors, until the moment at which any part of the vehicle reaches one of the lane boundaries. These periods of error-neglecting alternate with periods of error-correcting. Drivers' steering performance will be affected by inaccuracies in both their anticipatory and compensatory steering actions in curve driving. Drivers appear to be very well able to take account of curvature and speed effects when generating the anticipatory steering action and deciding on an appropriate speed at curve entrance. Research established that the reversal point on the inside of a curve, where the edge of the road reverses direction, serves as an important driver’s fixation before the vehicle enters the bend, and the driver’s gaze remains there for an appreciable percentage of time while traversing the curve. Furthermore, peripheral vision enables learning an automated routine of keeping the car in a lane. This parallel lane-keeping ability of experienced drivers releases mental resources for obstacle detection and other driving subtasks.

A.4.2 Vehicle control in straight lane keeping

Previously, most automobile steering control models were based on the assumption that the driver acts as an error-correcting mechanism with continuous attention allocated to the steering task. Godthelp et al. (1984) argued that driving cannot simply be considered as such a permanent closed-loop task. Instead of performing in a closed-loop mode, drivers may switch temporarily to visual open-loop control. Furthermore, their error-correcting behaviour may alternate with periods of error-neglecting. It can be argued that the time available for a driver to control the vehicle without immediate visual feedback, i.e. open loop, ultimately will depend on:

a) The accuracy of open-loop-generated steering actions.

b) The time available for error-neglecting.

According to Godthelp (1988), error-neglecting descriptions of driving should meet two major requirements: (i) it should give insight into the opportunity for error-neglecting at each moment of a run, i.e. independent of the instantaneous vehicle position, and (ii) it should allow for a quantification of the actual limitations of error-neglecting, i.e. the driver’s decision-making process.
in switching from error-neglecting to error-correcting, when approaching the edge of the driving lane.

A description which meets these requirements was developed by Godthelp and Konings (1981), by application of the path-prediction techniques commonly used in preview-predictor models. For each moment, the future vehicle path is predicted on the assumption that the steering wheel will remain in its momentary position during the time span of the prediction process. From such predictions time-to-line-crossing (TLC) can be calculated, representing the time a driver has available to neglect path errors, until the moment at which any part of the vehicle reaches one of the lane boundaries.

An experiment presented by Godthelp (1988) was designed to provide the rules for a straight lane-keeping task. Drivers were instructed to neglect the vehicle path error and switch to error-correcting only at that moment in time when the vehicle motion still could be comfortably corrected to prevent a crossing of the lane boundary.

A distinction was made between two hypothetical decision rules for steering control, A and B, as these may be adopted by drivers when deciding to switch from error-neglecting to error-correcting. In case A, drivers were assumed to use a constant lateral distance to the lane boundary for their decision to switch to error-correcting. A consequence of this rule would be that TLC’s, i.e. TLC at the moment of switching to error-correcting, are shorter with higher lateral speed. With decision rule B, drivers are assumed to compensate for higher lateral approach speeds by switching to error-correcting at a greater lateral distance from the lane boundary. A consequence of the latter rule might be that TLC’s are about constant for different lateral speed levels. The results of the empirical study indicated that drivers choose a strategy which corresponds closely with rule B - lateral distance from the lane boundary at which drivers switch from error-neglecting to error-correcting increased about linearly with lateral speed, whereas the TLC, values decreased only slightly with higher lateral approach speeds.

This finding can largely be explained by the tendency to choose larger distances $y_s$, with higher lateral approach speeds. Together, the mechanisms described result in an approximately constant TLC level for the range of speeds considered (20-100 km/h). In other words, drivers switch to error-correcting at roughly a constant time distance before the lane boundary would have been reached. This finding can be compared with results reported by Godthelp et al. (1984) on the duration of drivers’ self-chosen visual occlusion periods during straight lane keeping. They found
that drivers terminate occlusions so as to leave a constant fraction of the time available before the lane boundary would have been reached; this finding also being consistent over a broad range of speeds.

Together, these findings on visual open-loop control and error-neglecting indicate that the timing processes involved in these two steering strategies may differ fundamentally (Godthelp, 1988). During open-loop control, drivers have to rely on their estimate of the vehicle trajectory and this uncertainty results in the strategy of leaving a fraction of the time available at the end of the occlusion interval. In an error-neglecting task, with deliberate neglecting of path errors, drivers will be quite certain about the vehicle motion in relation to the lane boundary and this certainty results in the strategy of leaving a constant amount of time before the lane boundary would have been reached. These findings indicate that a driver’s method of timing, i.e. leaving a fraction of time or a constant amount of time, is strongly related to the degree of uncertainty about the vehicle trajectory.

The lateral distance from the lane boundary at which drivers switch from error-neglecting to error-correcting shows a smaller variability for leftward approaches (right hand driving). This tendency is most probably related to the driver’s offset position in the car, which allows for a better observation of the distance to the centre line than to the edge line of the road.

A.4.3 Vehicle control during curve driving

TLC analysis has been used for straight lane keeping. Godthelp (1986) also applied the approach to describe vehicle control during curve driving. Previous steering-control models for curved-road driving presented by Allen and McRuer (1977), and Donges (1978) assumed an error-correction or compensatory mode to function in parallel with a so-called anticipatory mode. The anticipation mode generates steering actions, \( \delta_{sa} \), using the previewed road curvature as major input and translates this curvature into a steering-wheel movement. On the basis of instantaneous perceived path errors, the compensatory mode generates correcting steering-wheel movements, \( \delta_{sc} \), which are taken together with \( \delta_{sa} \), resulting in an overall steering \( \delta_s \).

Crossman and Szostak (1968) indicated that the anticipatory and compensatory modes should be considered as acting in serial order rather than in parallel. They argued that, particularly at curve entrance, steering will be based primarily on the anticipatory mode, whereas compensatory control comes into operation only after this initial steering action. Although this reasoning intuitively
seems correct, the closed-loop situation hardly permits any verifiable distinction between these modes. Therefore, a temporary withdrawal of visual feedback during the curve entrance phase was used by Godthelp to force the driver to rely only on the anticipatory mode during a certain time period. Then, by analysing the accuracy of the anticipatory steering action, it became possible to quantify the potential role of error neglecting immediately after this action; that is, during actual curve negotiation.

The anticipatory steering action can be regarded as the outcome of an information-processing chain that contains three major stages: Stage A, perception of the curve, resulting in an estimated curvature; Stage B, translation from estimated curvature into a desired steering-wheel position; and Stage C, motor control process to transform the desired steering position into manual action.

Drivers’ steering performance will be affected by inaccuracies in each of these stages. Godthelp (1985) has analysed accuracy in precognitive steering tasks previously. In experiments on lane changing, the accuracy of the steering-wheel angle amplitude was measured for the pull-out steering action in a lane-change manoeuvre, under conditions with and without visual occlusion. By having subjects repeat the same manoeuvre 36 times in succession, a quantification could be made of the ultimate steering accuracy in such a pre-programmed task. The inaccuracies measured in such a pre-programmed task are assumed to originate mainly from the aforementioned Stage C. In a preview control task like entering a curve, the inaccuracies of the sub processes A and B can be expected to be added to those of the motor-control process.

Drivers appeared to be very well able to take account of curvature and speed effects when generating the anticipatory steering action at curve entrance. This important finding illustrated that experienced drivers have a rather good internal representation of the vehicle characteristics. By consequence, it may be expected that drivers use this knowledge during curve negotiation to rely on internal feedback processes, thus allowing a temporary loss of external, visual feedback. The limitations of this feedback process appear to be strongly related to the inaccuracies of the anticipatory steering action. For both the conditions, with and without occlusion, these inaccuracies increase with the extent of the anticipatory steering-wheel angle and thus with road curvature.

Godthelp (1986) stated that the necessity for compensatory control actions after the anticipatory steering-wheel movement would be strongest for sharp curves; that is, those requiring large steering-wheel amplitudes. This assumption was confirmed both in terms of the steering-wheel
velocity data and in the TLC analysis which showed TLCs to be shortest for the sharpest curves. This latter finding clearly illustrates that in sharper curves, drivers will have to switch over to the error-correcting mode at an earlier moment in time. The fact that the results are found both in conditions with and without immediate visual feedback supports the assumption that the anticipatory steering mode plays a dominant role during the curve entrance phase.

A.4.4 Choice of speed and steering behaviour during curve driving

Car-driving behaviour in curves may be regarded as an interesting case in which steering, as an example of operational performance, is intimately related to behaviour on the tactical level – in this case, the choice of speed as a function of curve radius. The distinction between the operational and the tactical level of car-driving behaviour has been made by several authors (Michon, 1985). Until now, studies of car-driving behaviour in curves have focused exclusively on either speed choice or steering behaviour, and no attempt has been made to integrate these two lines of research (Van Winsum and Godthelp, 1996). Results have encouraged the idea that drivers use lateral acceleration as a cue in speed choice, in that they accept a smaller lateral acceleration as a safety margin at higher speeds (and thus larger radii). Curve radius and speed during curve negotiation affect required operational performance because both factors affect the required steering wheel angle.

If steering wheel angle during curve negotiation matches the required steering wheel angle perfectly, speed is restricted only by an upper limit at which the vehicle begins to skid. There is some evidence that steering errors increase linearly with required steering wheel angle (Godthelp, 1985, 1986). There is also evidence that steering error is affected by steering competence. Cavallo et al. (1988) found that under visual occlusion, experienced drivers estimated the correct required steering wheel angle better than did inexperienced drivers.

According to Van Winsum and Godthelp, the model of the relation between speed choice and steering performance may be summarised as follows: Required steering wheel angle is determined by curve radius and speed, whereas steering error is determined by required steering wheel angle and steering competence. It is assumed that the driver has learned the effect of curve radius and speed on required steering wheel angle and on steering error from previous experience. In addition, it is assumed that steering error is consistent and that the driver is aware of his/her steering competence.
When the driver approaches a curve, both radius and steering competence cause an anticipatory adjustment of speed. The effects of radius and steering competence on steering error are traded off with speed, such that the safety margin TLC remains constant and independent of radius and steering competence. The relation between lateral acceleration and speed is then assumed to be a by-product of this mechanism.

Van Winsum and Godthelp found experimentally that both required steering wheel angle and steering error during the open- and closed-loop phases increase with smaller radii but that the relative steering error, defined as steering error divided by required steering wheel angle, is constant over radii. Smaller radii resulted in the choice of a lower speed, but the minimum TLCs during curve negotiation were not affected by radius. The TLC as a safety margin, then, is controlled by the driver's speed choice. The results suggest that speed choice and steering performance are both intimately related in negotiating curves.

Because steering competence did not affect TLC, it cannot be concluded that drivers with poorer steering competence were less safe drivers.

It thus appears that both radius as a road design element and steering competence as a driver characteristic exercise their influence on driving behaviour in the same manner. Both affect operational performance, resulting in an adaptation of behaviour at the tactical level in an attempt to control safety margins.

A.4.5 Fixation during car driving

Steering a car requires visual information from the changing pattern of the road ahead. However, there is little direct information linking steering performance to the driver's direction of gaze. Land and Lee (1994) have made simultaneous recordings of steering-wheel angle and drivers' gaze direction during a series of drives along a tortuous road.

Viewing the videos of the three drivers tested, they were immediately struck by the way the eye seeks and returns to the tangent point (reversal or extremal point) on the inside of each bend or curve, where the edge of the road reverses direction. The drivers themselves were surprised by their consistent use of this point. Gaze was directed to the tangent point, with a rather obvious saccade 1-2 sec before the car entered each bend, and remained there with relatively few excursions...
for about 3 sec into the bend. In addition, half a second after the car has entered the bend, the gaze of the drivers was directed to the tangent point for about 80% of the time.

On a narrow undulating road, the tangent point is almost the only reliable cue to curvature because, unlike the general shape of the road ahead, it is unaffected by slope changes. It is possible to steer while deliberately looking at the opposite side of the road, but it feels wrong, and one has to drive more slowly. Here, as in other tasks, the brain prefers a ‘do it where you look’ strategy, in which the object guiding a control strategy is placed close to the centre of vision, according to Land and Lee. In more relaxed wide-road driving, much more time is spent on tasks irrelevant to steering, which is why the importance of the tangent point has not previously become apparent. In urban driving, where one may be called upon simultaneously to steer a course, avoid other vehicles and look for a street name, this capacity for dividing time while avoiding cross-talk between tasks is crucial.

A.4.6 Maintaining lane position with peripheral vision during in-vehicle tasks

An old tenet in the traffic safety community states that drivers learn to use peripheral vision in lane keeping, whereas beginners need focal vision for it - this was hypothesised by Mourant and Rockwell (1972) on the basis of their eye movement studies. They showed that novices’ fixations scatter more and also cover lane boundaries, whereas with increasing practice, drivers’ fixations focus more at the vanishing point on the horizon, at a greater distance from the vehicle.

Nevertheless, there are obvious limits in the use of peripheral vision. Hella (1987) showed that although drivers are able to control lateral position while reading a display mounted in the parafoveal field of vision, at the lower edge of the windshield, or at speedometer level, the lateral scatter starts growing when the display is mounted at the top of the windshield or down in the middle console. In the case of obstacle avoidance, and even movement detection, thresholds and response latencies also grow with the eccentricity of stimuli because of the inferior performance of peripheral vision (Saarinen, 1993); and because of the eye and head movements required at large eccentricities (Sanders, 1963).

Furthermore, people’s attention capacity is both limited and task dependent. The useful visual field gets narrower with increasing mental load (Ball and Owsley, 1991), and attention sharing in a dual-task paradigm is dependent on the type of resources required for each task (Burris et al., 1994). In studying driver performance during in-car tasks, therefore, one should take into account both the
position (eccentricity, or the visual angle from the normal sight axis) and the mental load of the task.

In an experiment by Summala et al. (1996), the attention tasks showed a consistent interaction between task position and driving experience. Novices' lane-keeping performance was already deteriorating with the task at the level of the speedometer (23° from the normal sight axis), whereas experienced drivers' performance was impaired only when the task was low on the middle console (38° from the normal sight axis). Results further suggested that the superior 'peripheral' lane-keeping ability of experienced participants was based on an automated, parallel routine rather than on serial attention sharing in the visual field.

Considering the experience effects, however, it is not sufficient to speak only of learning to use peripheral vision rather than foveal vision in lane keeping. Practice at the wheel means improved use of visual information, including the more efficient use of different cues, whether they be based on foveal or peripheral information. Thus in lane keeping, Smiley et al. (1980) found an early change in the steering control strategy among beginners during the first days of driving, and Riemersma (1987) found that experienced drivers use lateral speed as an effective visual cue for straight lane keeping but that inexperienced drivers do not. Therefore, more efficient cues learned with practice make the use of peripheral vision easier either because the task gets generally more simple or because the cues are better suited to the periphery.

Although a distinction is traditionally made between focal and ambient processing in the visual system, the results from Summala et al. indicate that the use of peripheral vision in driving can be learned through practice (also Engel, 1971; Johnson and Leibowitz, 1975). What is more, practice is presumably required for parallel, ambient processing. Conversely, the results support the view that driving is learned by driving. Learning an automated (more peripheral) routine of keeping the car in a lane, which probably coincides with (or incorporates) automatisation of steering control, means a profound change in the entire task. Parallel lane keeping releases resources for obstacle detection and other driving subtasks (Summala, 1996), and although overt behaviour may remain the same – at least in a supervised experiment – an experienced driver with his/her expertise is actually performing another task and is doing it more safely.
A.4.7 Application of approximate time-to-line-crossing (TLC) during car driving

Incidents and accidents in which the driver exceeds the lane boundaries or inadvertently moves off the road are often preceded by a period during which the TLC minima are already low (Van Winsum et al., 1999). This suggests that control of lane position diminishes as drowsiness progresses. TLC minima may then indicate progressing drowsiness and be used to warn the driver of deteriorating performance before the vehicle actually drifts out of the lane. Another important application of TLC in driver warning systems is to detect instances when the vehicle actually moves out of the lane and to warn the driver in order to avoid an immediate accident. Both application of TLC are useful in driver support systems to warn the driver in case of severe impairment caused by drowsiness. The second case, i.e. drifting out of lane, indicates severe impairment where the driver actually falls asleep or loses control of the vehicle, while the first case would suggest a milder form of impairment and warnings would be given at an earlier stage. However, the use of TLC in both cases is different. In the first case, TLC minima are used to evaluate driver performance. A minimum occurs in the TLC signal only if a steering correction has been initiated by the driver. If a minimum occurs, the lane boundary will not be exceeded. A TLC minimum will not occur prior to exceeding a lane boundary. Here, the TLC decreases until it passes through zero when the lane boundary is actually exceeded.

A major problem with the measurement of TLC is the complexity of its computation in real time while driving on the road. The essence of TLC calculation is that the vehicle is rarely driving according to a straight path: the movement of the vehicle can be described as a curve most of the time in which curves to the right are alternated with curves to the left. In experimental practice, TLC is often computed as lateral distance divided by lateral velocity. This would only give reasonable results if lateral velocity, i.e. the velocity at which the lane boundary is approached, were constant. Since vehicle movements generally follow a curved path, lateral velocity is rarely constant and this popular approximation of TLC would give highly inaccurate results. Therefore, the accuracy of approximations of TLC depends on both the specific manoeuvre performed by the driver and the purpose for which TLC is used.
A.5 Drivers’ Visual Search

A.5.1 Summary

Research covers the questions of how drivers distribute visual attention and on the differentiation of observed objects into classes. In order to assess the fraction of time directed to driving-related tasks, such as searching for needed visual information relevant to vehicle control and guidance, researchers have to go to great lengths to ensure the utility and validity of verbal reports from drivers. Roadway advertising seems to simply attract spare attention drivers have to sample the visual environment for objects unrelated to driving, rather than from diverting attention from those thought to be relevant to the driving task. However, there is a trade-off with evidence that visual clutter reduces the conspicuity of objects, including traffic control devices and other traffic. Eye-head dynamics impose a substantial risk level on certain driving manoeuvres in that visual input is completely removed from critical inputs, incurring periods of not tracking important traffic dynamics, such as a rear or mirror search during car following. Co-ordinating visual input with control movements is an element of the driving task which a novice driver must master. This is apparent from the differences in visual search strategy between novice and experienced drivers, such as fixation durations and the spread of search along the horizontal and vertical axes for various road environments. The control of visual sampling is described as visual attentional switching when a mental model is triggered by salient dynamic information, while the mental model in its turn samples dynamic information and construct a ‘calling priority’ in the brain when anything changes in a noticeable way.

Deteriorating eye health (central and peripheral vision) and visual function together with mental status are now believed to impact negatively on the useful field of view of older drivers and their accident proneness. Results support the hypothesis that detecting closing-headway situations with peripheral vision does not improve with driving experience at the same rate as lane keeping does. Together with the fact that an in-car task also causes the two driving subtasks to deteriorate differently, can therefore mislead a driver with accumulating experience. Early experimental work has shown that perceptual and decision-making tasks are impaired when a driver has to divide attention between the road and a car phone. However, drivers also feel that mobile phones in cars in some instances increase traffic safety. Research also shows a smaller variance in the length of glances at the secondary task (in-car task) among experienced drivers than inexperienced ones. Over-long glances of novices are a potential safety risk as they are more prone to extreme lateral displacements increasing the chances of running off the road.
A.5.2 What attracts attention when driving? Drivers report

Conspicuous targets are those that have a high probability of being seen within a very short observation time. Pattern research has shown that several physical parameters are important determinants of conspicuity, notably eccentricity, background complexity, contrast, colour and the boldness of the internal structure of the object. However, cognitive factors are also important and two measures of conspicuity, attention conspicuity and search conspicuity are proposed to account for two basic situations, one when the observer is actively searching for needed visual information and the other when attention needs to be attracted to unexpected information. A field trial using these concepts has shown that drivers mostly notice objects or find them by search when they are located within +/- 8 deg of the direction of travel of the vehicle (Cole and Hughes, 1988).

For the traffic engineer faced with the responsibility of ensuring that drivers notice information displays relevant to vehicle control and guidance, the first need is a body of data describing how a driver distributes attention and what classes of objects attract attention. In a paper by Hughes and Cole (1986), the results of experiments to provide some information on the topic, is reported.

Ericsson and Simon (1980) argue at length that verbal reports can be regarded as valid data and they present a model of how subjects, when asked to think aloud, verbalise information that they are attending to in short-term memory. They show that the process of verbalising will affect the cognitive processes only if the instruction requires verbalisation of information that would not otherwise have been heeded or attended to. Thus verbal reports can be used to identify the information that is attended to in short-term memory provided the instruction does not require a selective response, since this must generate an additional cognitive process to test whether the heeded information matches the selected type. A further requirement, if verbal reports are to reflect attentive processes validly, is that instruction should not require description by the subject of his/her inferences and motor actions, since this would require the subject to heed his/her own internal processes in addition to the information in short-term memory.

It is then possible to probe the processes of attention when driving without disturbing the usual cognitive processes by simply asking subjects to report what attracts their attention without any further or more specific instruction. Since the capacity of short-term memory is limited to a small number of chunks of information, and stored information is obliterated by new incoming information, only the most recently heeded information will be accessible to verbalisation.
Ericsson and Simon (1980) also argue that verbal reports of heeded information in short-term memory will tend to be complete provided the subject is not under high cognitive load since major task-directed processes will tend to take priority over the process of verbalisation. There may also be failures to verbalise heeded information when the processes are highly practised and have become automatic. They argue that highly practised processes become more and more automated so that information of the steps in the processes are not held in short-term memory.

Therefore, it is possible that the task of driving may impose a high cognitive load and cause incomplete verbalisation and it is also possible that some processes involved in driving are so automatic that there may be no intermediate stage of recognition of information in short-term memory.

In order to test the effect of the driving task on the pattern of verbal reports Hughes and Cole conducted two experiments, one with the subjects driving a motor vehicle and the other with subjects simply observing a movie of the route driven in the first experiment. This comparison would test whether driving imposed a significant cognitive load taking priority over verbal reporting and whether the automatic processes of driving lead to some elements of information not reaching short-term memory. The comparison would also test whether investigations of attentive behaviour could be pursued in the laboratory, which would be a useful result since field investigations are time-consuming and pose difficulties of experimental control.

A total of 7721 reports were made by the 50 observers. These were classified into eight categories which were selected on the basis of the experience of Renge (1980) and Lynch and Rivkin (1959) with verbal report studies of a similar kind. The categories were:

a) Road related: reports of road geometry and road surface including surface delineation marking and pavement messages.

b) Traffic control devices: reports of any road-traffic control device including traffic signs, traffic signals and road delineation chevrons.

c) Vehicles: reports of moving and parked vehicles including reports of vehicle signal lights.

d) People: reports of people mostly of pedestrians on the footpath and roadway.

e) Immediate road surround: reports of objects located in the gutter or between the gutter and the building line. These reports consisted mainly of rubbish bins, letterboxes and objects related to service reticulation.
f) General surrounds: reports of objects beyond the building line but not including reports of vegetation. These reports were mainly of buildings, houses and shops.

g) Vegetation: reports of any kind of vegetation, but most reports were of trees and shrubs.

h) Advertising: reports of any kind of advertising including signs, neon displays and billboards.

Laboratory replication of visual behaviour in the field

Their results indicated that a movie film provides a reasonable good simulation for studies of the distribution of attention using a concurrent verbal report technique. Although there was a loss of high spatial frequency information with movie film, and contrast and colour rendition could not be exact, the reporting behaviour of the subjects who observed the film was very similar to that of the subjects who participated in the field trial. This then confirms the conclusion reached by Macdonald and Hoffmann (1984) who found that retrospective verbal reporting of road sign information in a laboratory experiment using movie film satisfactorily replicated results obtained in a field trial. The study by Hughes and Cole also extends that conclusion by showing that a laboratory study using movie film can yield essentially the same pattern of reporting as a field trial for all classes of objects, not just road signs.

The effect of driving on visual behaviour

The task of driving was not simulated in the laboratory trial - the subjects simply watched the movie film and reported what attracted their attention. However, the absence of the driving task did not have any substantial effect on reporting behaviour. It might have been expected that the laboratory observers would have noticed and reported more objects, since they were free of the demands of vehicle control, and that a greater proportion of their reports might have been unrelated to driving since they had no need to make decisions about driving.

There were more reports from the laboratory observers than there were from the field trial observers but the difference was small and not statistically significant. The pattern of reporting was very similar in the two experiments and the attentive behaviour of the laboratory trial observers did not in any way reflect the fact that they were free of the decision-making responsibility of driving. However, they made significantly more reports of traffic control devices and of vehicles than did the field trial observers. According to Hughes and Cole, the field trial observers made fewer reports of these classes of driving-related objects because intervening task-related cognitive
processes took priority on occasion and displaced the verbal reporting process. Alternatively, it may have been because the decision-making process while driving was sufficiently automatic on occasions for the initiating stimulus not to reach short-term memory.

If this is so, it can be argued that the reports made in the laboratory trial are more complete and better reflect the attentive processes of the driver than do the field trial reports. Support for this contention lies in the fact that the reporting patterns for the reports not related to driving were almost identical for field and laboratory trials which might suggest that some reports of driving-related objects were lost in the field trial because of the operation of automatic processes. Despite these differences, the results strongly suggest that the laboratory observers behaved like drivers even though they did not have the responsibility of vehicle control and were not subject to dual task demand. Perhaps this is not surprising since motor vehicle passengers do often behave as ‘back seat drivers’.

Attraction of attention by advertising

According to Hughes and Cole, it is tempting to note from their results that when advertising occurs it occupies a significant proportion of drivers’ attention and from this to infer that if there was less advertising a greater proportion of attention would have been directed to driving-related tasks and that as a result a greater number of traffic control devices would have been noticed.

In one sense this may be a correct inference. There is converging evidence that visual clutter reduces the conspicuity of objects (Hughes and Cole, 1984; Macdonald and Hoffmann, 1984) and the removal of advertising would reduce visual clutter, thereby rendering other objects more conspicuous.

However, the increased attention given to advertising in arterial and shopping centre roads was offset by decreased attention given to other objects unrelated to the driving task and that the proportion of attention given to objects unrelated to driving actually decreased. The 30-40% of attention given to these objects may represent the result of the spare capacity the observers had to sample the visual environment and that the reports made are of the objects that adventitiously or accidentally attracted attention during the sampling process. Thus if advertising signs were limited or removed the reports may still be of objects unrelated to the driving task. In the residential sections of the route, there was little advertising but some 50% of reports were of objects unrelated to driving and were of such inconsequential objects as garbage containers, letterboxes, bald joggers
and pregnant pedestrians. Therefore, removal of advertising in the shopping centres would simply divert attention to objects of this kind and not to those thought to be relevant to the driving task.

A.5.3 Eye-head dynamics, visual search and risk levels

Driving an automobile requires continual and complex visual search. Traffic, pedestrians, road signs and a multitude of other possible objects provide a rich visual input which must be processed quickly and in proper sequence if the task is to be performed efficiently and safely.

Robinson et al. (1972) presented the results of two experiments designed to study the visual search of drivers during two critical driving manoeuvres: entering a highway after a stop, and changing lanes on a multilane highway. Both of these manoeuvres place severe demands on the driver’s visual input system, primarily because of the large visual angles between the necessary information sources.

Eye-head dynamics

The data in both experiments were obtained measuring head movements only. The actual visual search pattern is, of course, the result of both eye and head movements. To facilitate the interpretation of the head movement data in terms of actual visual fixation a series of laboratory measurements were also taken in which both eye-movement-with-respect-to-head and head movements were measured. Subjects were commanded to refixate, as quickly as possible, on targets at angles of 40°, 60° and 80° from a resting point straight ahead. The findings of importance to the interpretation of the highway data were that: (i) the dynamic relationship between the eye and the head is reasonably stable for an individual, and rather remarkably consistent between individuals, (ii) the head initially lags the eye by about 50 msec for all target angles, (iii) eye movement stops at about 40° for all target angles (greater than this value), and (iv) the maximum velocities for the eye are on the order of 1 000°/sec, and for the head about 450°/sec.

The implications of these data for the highway experiments under consideration are that: (i) head-movement data are sufficiently accurate for practical measurements where refixations are greater than 20° or 30°; and (ii) the actual visual fixation time is approximately 50 msec longer than that indicated by the head ‘fixation’ data, due to the higher velocity of the eye. This gain in fixation time is reasonably independent of target location for angles greater than 40° since the gain made by the eye ceases at its maximum excursion.
Search times, search patterns and task complexity

Individual visual search times in Experiment 1 (stop and enter) ranged from 1.1 to 2.6 sec. In Experiment 2 (lane change) search times ranged from 0.8 to 1.6 sec for the mirrors and 0.8 to 1.0 sec for search back (including movement times). The longer times in Experiment 1 are probably a function of two conditions: the static environment of Experiment 1 wherein long search times are not unsafe; and the probability that in Experiment 1 the driver was accumulating more information in each search. In Experiment 2, the driver was probably limiting himself to a yes-no decision. However, in Experiment 2 mirror times are long, particularly for those drivers making almost exclusive use of the mirrors for information.

Increased complexity of the visual input task (for example, more traffic) leads primarily to more visual searches, and, to a lesser extent, to longer individual searches. This, of course, is an optimal strategy, given the increased risks incurred during periods of not looking at critical inputs, such as a lead vehicle in Experiment 2.

Risk levels imposed by necessary search

In both of the driving manoeuvres, a substantial risk level is imposed by the limitations of driver search. The primary risk in the manoeuvre of Experiment 1 (stop and enter) is what might be called the ‘last-look’ problem. If, on entering a highway, the last visual search is to the right, then unpredicted changes on the left, occurring during the search right, will be undetected, at least until the actual entry is begun. It appears that about 1 sec is the minimum loss that can be achieved and therefore any traffic or pedestrian dynamics within one second may be critical.

For Experiment 2 (lane change) the high-risk problem is search to the rear, a search element used in almost all of the observed lane changes. As in the right-left entry problem (Experiment 1), the rear search removes the visual input completely from important other sources. The time appears to have a minimum value of about 0.8 sec (including movement) and thus any traffic dynamics ahead of the vehicle that occur within 0.8 sec (plus brake or steering reaction time) may be critical. Freeway traffic has been frequently observed with gaps of one second or less, indicating rather clearly the possible causative role of visual search in freeway rear-end collisions.
A.5.4 Strategies of visual search by drivers – fixation characteristics

Coordinating visual input with control movements is an element of the driving task which the novice driver must master. Fitts and Posner (1967) listed three stages in learning skilled tasks:

a) Early or cognitive phase: At this stage, performance is simply a patchwork of old habits and looks awkward and discontinuous.

b) Intermediate, or associative phase: During this phase, new patterns of individual unit of behaviour emerge and errors are gradually eliminated.

c) Final or autonomous phase: During this phase, performance is inflexible and automatic. Many visual cues attended to in the early phase now appear to be redundant.

In an experiment by Mourant and Rockwell (1972), the median horizontal and vertical fixation locations indicated that the central direction of gaze of the novice drivers was lower and farther to the right than that of the experienced group (for right hand driving). It appears that this location was due to the novice drivers’ frequent sampling of the curb in order to verify or estimate vehicle lane alignment. In a study of driving fatigue, Kaluger and Smith (1970) also found that drivers looked more to the right and closer in front of the vehicle after several hours of driving. The compensating strategy of fatigued drivers appears to be a regression towards the visual scan behaviour of the novice drivers in neighbourhood driving.

Furthermore, Kaluger and Smith (1970) found that both the duration and frequency of pursuit movements increased with amount of sleep deprivation. Kaluger and Smith reanalysed Zell’s (1969) data and found that novice drivers also made pursuit eye movements. Because of the motion of the vehicle, pursuit eye movements are related to how closely the driver is sampling the area in front of the car using foveal vision. Thus, pursuit eye movements on lane markings may indicate the times when a driver is critically concerned with the vehicle’s lane position and direction of travel.

Learning to sample the driving environment visually is characterised by a long recognition stage as compared to other skill learning. This may be due to several factors; one being that the many different types of information that a driver must sample are only available for brief periods because of the motion of the vehicle. A driver must make many fixations on a stimulus pattern, at discontinuous points in time, in order to learn which features are redundant. Another factor may be that the novice driver’s primary task appears to be maintaining control over the vehicle. Further, the psychomotor feedback loops of relating vehicle changes in direction and velocity to control
movements may take more time to develop into perceptual reflexive responses than is generally realised. A third interpretation is that the long recognition stage may be due to the complex interaction between foveal and peripheral vision that is demanded by automobile driving. A novice driver samples considerable information foveally which the experienced driver obtains through peripheral vision. Thus, novice drivers must learn to recognise objects and relationships with peripheral vision (Mourant and Rockwell, 1972).

A.5.5 Strategies of visual search by drivers – recent findings

The analyses (Crundall and Underwood, 1998) of mean fixation durations and the spread of search along the two axes suggest that there is an influence of experience on the effects of processing demands in driving. With regard to mean fixation durations, the reported finding that novices produce longer fixation durations than experienced drivers (Mourant and Rockwell, 1972) is not a simple difference but one which depends on the type of road being driven on at the time. If one views pursuit tracking as the foveation of a moving object (as the viewed image is fixated in the sense that it is held in place on the fovea), the increase in these movements may also be explained in terms of the extra processing time required by novices due to the novelty of the stimuli. Both the experienced and novice drivers displayed sensitivity to the different road types in their fixation durations though their responses tended to opposite directions. If the rural road is viewed as the least demanding due to the low levels of traffic, lack of parked vehicles and pedestrians, and general absence of visual complexity, then the experienced drivers seemed to increase their fixation durations on the least demanding of the roads whilst novices increased fixation durations on the more demanding dual carriageway. The only roadway where the drivers distribute their visual attention in similar ways is the suburban route.

If one considers traditional research findings in the areas of reading or picture viewing, increased fixations are interpreted as extra processing time due to a complex foveal stimulus (Mackworth, 1976; Loftus and Mackworth, 1978; Henderson et al., 1987; Underwood and Everatt, 1992). On the dual carriageway the novice’s behaviour reflects this. The experienced drivers, however, have found the dual carriageway and the suburb to be the least demanding by this analogy. One could also posit that the reduced durations may be part of a compensation strategy to deal with the increased demand (Miura, 1990) by reducing the time spent foveating any one location so as to sample more of the scene on the complex roads, a strategy which the novices have yet to develop on the dual carriageway. Consistent with this explanation is the result that the suburban route produced the most fixations.
Crundall and Underwood's results are similar to the findings of other studies which have also shown decreased eye fixation durations when driving through increasingly demanding roadways. Several researchers have noted a decrease in fixation duration and an increase in the number of fixations when driving through a curve compared with a straight roadway (Shinar et al., 1978; Zwahlen, 1993), and others have noticed an inverse correlation between fixation duration and speed (Cohen, 1981).

The notion of a compensation strategy is supported by their finding that experienced drivers also increased their visual search in both the horizontal and vertical planes on the dual carriageway. Along both axes the experienced drivers gave an extremely wide search on the dual carriageway relative to their search on the other roads. The rural road produced the narrowest search in experienced drivers along the horizontal meridian, which may reflect the length of the fixations that this road evoked. Thus, the theoretically least demanding roadway produced an extremely small search strategy which increased as the situation became more demanding.

The novice drivers did not vary the size of their visual searches along either the horizontal or vertical across the three roadways (i.e. dual carriageway, rural road, and suburban route). They produced a similar horizontal search to that of the experienced drivers on the suburban and rural roads, while their vertical search resembled the experienced drivers search strategy on the dual carriageway. The novices adopted the restricted search of the rural and suburb roads for the dual carriageway, and conversely imposed the vertical search strategy of the dual carriageway upon the rural and suburban routes. Regardless of the level of demand that they pitched their strategies at, it is important that they did not vary this strategy across the three roads. In this sense, their strategies seem to be too rigid to be adapted to the differing demands of the roads.

The description of events suggests that novices may not have developed the flexible approach to viewing dynamic scenes required in the real world of driving. Instead of expanding their visual search to cover demanding roads they restrict it and concentrate on the area of the scene that proves the most novel, dangerous or simply hard to process. They basically take the search strategy of the pedestrian or static picture viewer, and transfer this to the driving situation. When reading text, viewing a picture, or walking down a street, concentrating attention on particular stimuli may be a valid strategy to process them, though when travelling at more than 46 km/h any form of attentional capture could be dangerous. The novice's longer fixation durations on the dual carriageway may represent a form of attentional capture. Traffic in front, travelling in the same direction, grabs
attention as it is the most likely source of any requirement for evasive manoeuvres. This situation might arise more on the dual carriageway than on the rural road, if not the suburban route also. Perceptual narrowing may also play a role. If the dual carriageway is highly demanding then the peripheral field may shrink. The reduction of peripheral and parafoveal preview would also increase fixation durations (Henderson et al., 1989).

Another example of novices who have not yet learned the optimum search strategy for driving was the consistent difference reported between novices and experienced drivers with regard to vertical search, with novices producing a wider search than experienced drivers seem to deem necessary (Renge, 1980; Evans, 1991). It has been suggested that this reflects novices’ lack of sensitisation to the horizontal axis as the major source of information. However, Crundall and Underwood (1998) have shown that under certain circumstances, experienced drivers may increase their vertical search to a level comparable with that of novices, and thus it would be a mistake to condemn an expanded vertical search strategy as the sole property of the inexperienced driver.

A.5.6 Reading the road - visual attention and saliency

According to Underwood (1998), there is a facile, but relatively meaningless analogy to be drawn between reading text and ‘reading the road’ – the phrase widely used to describe some interpretative advantage experienced drivers are assumed to have when handling the mass of visual information they confront when driving. With reading text it seems likely that a familiarity check is first carried out, which programmes the next saccades, with lexical access following and resulting in a shift to the next word when completed. When less plausible, or unfamiliar, stimuli are processed by parafoveal regions of the eye, this relatively smooth predictable process is disrupted, because the additional processing capabilities available for foveated stimuli are required.

Moving stimuli attract attention more than static stimuli. Where these are detected in areas adjacent to foveally fixated regions, the normal processes thought to underlie static scene perception are substantially changed. Otherwise, aspects of the scene are successively subjected to detailed analysis, driven by what is termed a ‘salience map’. First, visuo-spatial attention is allocated to the scene region with the highest saliency weight, with further attentional resource being deployed in order to keep the eyes fixated on the attended scene region. The eyes remain fixated on the selected element of the scene until perceptual and cognitive analysis is completed, or until the rate at which information is gained about that region slows markedly. In the first case, the saliency weight for that region is reduced and attention will be released and allocated to the region which
now has the highest saliency, and the eyes are programmed to move to that region. Where analysis of the current region is slow, perhaps because of a scene’s complexity, a within region re-fixation will occur, in order to acquire additional information, or simply because of occulo-motor factors. As individual scene regions are fixated and cognitively analysed, saliency weights will be modified to reflect the relative cognitive interest of those regions.

Where one looks in real scenes will thus be determined by why one is looking. It will also depend on the experience one has of looking successfully for certain types of things, which is what Underwood believes underlies the construction of the ‘salience map’ of a scene. Searching for and fixating a traffic sign will result in a different scanning pattern than when a driver is looking at the focus of expansion, the point of tangent on a curve, or the dashboard of the car he/she is driving.

A.5.7 Models of visual attention

In tasks involving human behaviour in transportation, information acquisition is largely dominated by the limitations of eye movements.

Information can be acquired from the periphery of vision, particularly about rapidly changing signals, and when the information displayed in the periphery is redundant with other sources. But essentially vision is a system which samples fewer than three points in space per second. This severely limits the amount of information which is available to the observer.

Given such an impoverished sampling of the environment, and looking at the rates at which people operate transport systems, how do they manage at all? The answer is three-fold (Moray, 1990). In the first place, the environment is highly structured. A glance at one part of the world tells us not merely what its state is, but also a great deal about neighbouring regions. The redundant structure of the world allows correlation to substitute for observation. Secondly, the bandwidth of the world is limited. In the case of transportation systems in particular, the bandwidth of the world is generated by the rate at which a vehicle traverses the environment. The faster one drives, the greater the rate of change of the visual array, and the greater the bandwidth of information generation. However, it always remains low compared with the bandwidth of the visual system. Thirdly, the paths taken by vehicles are severely constrained by physics. Physics constrains motion to rather predictable paths for most of the time. Hence, the motion of vehicles is to a great extent predictable.
There are now about 10 models of the control of visual sampling behaviour (Moray, 1987). What they all have in common can be summarised as follows. When a person has performed a task which involves visual monitoring and visual guidance for a long period (at least tens of hours) in an environment which can be regarded as statistically ergodic, he/she acquires a mental model of the dynamics in space and time of the variables relevant to the task and of the costs and rewards associated with different events and outcomes. This internal model of the environmental and task dynamics control the scanning pattern and rate at which sources of information in the environment are sampled.

Since the construction of the mental model is triggered by dynamic information, the model in its turn samples dynamic information, anything that changes in a noticeable way will construct a ‘calling priority’ in the head.

A.5.8 Visual attention problems of older drivers

Sensory tests, such as visual acuity and peripheral field sensitivity, although quite appropriate for the clinical diagnosis and assessment of ocular disease and vision loss, do not by themselves reflect the visual complexity of the driving task, and therefore would not be expected to reveal a strong relationship between vision and driving. The visual demands of driving are quite intricate. Controlling a vehicle takes place in a visually cluttered environment and involves the simultaneous use of central and peripheral vision and the execution of both primary and secondary visual tasks. The driver is usually uncertain as to when and where an important visual event will occur. Visual sensory tests do not typically incorporate these stimulus and task features, but instead seek to minimise perceptual and cognitive influences.

A test of visual attention consists of a central target identification task coupled with a peripheral target localisation task, which together provided a measure of the size of the useful field of view (UFOV) (Ball et al., 1988; Sekuler and Ball, 1986). The UFOV provides a measure of the spatial area within which a person can be alerted to visual stimuli in a variety of situations (Sanders, 1970; Verriest et al., 1983a; Verriest et al., 1983b). It is conceptually distinct from the visual sensory field that describes luminance sensitivity throughout the field (Ball and Beard, 1990). Preliminary work indicated that for all older observers without significant visual field loss, the peripheral localisation task could be performed with 100% accuracy at all eccentricities evaluated (10°, 20°, and 30° visual angle) without the central task and distracters. However, some older observers required longer stimulus durations than others to achieve this level of performance, which indicates
a relatively slower speed of visual processing. With the addition of the central task, the localisation performance of a subset of these observers was hampered, such that the most eccentric targets could no longer be localised (a reduction in the size of the UFOV). This reduction could, however, be counteracted by increasing stimulus duration to regain 100% accuracy in the dual task situation. Finally, the addition of distracters also impaired localisation performance for a subset of older observers, again such that the most eccentric targets could no longer be localised. This effect could sometimes be reversed by increasing target duration to compensate for the distracter effect. The size of the UFOV, defined as that eccentricity at which observers can localise peripheral targets correctly 50% of the time, is thus a dynamic measure. It is a function of (at least) three test variables: the duration of target presentation, the level of complexity of a secondary central task, and the salience of a peripheral target (where no distracters represents the optimal salience). Thus, the UFOV test incorporates stimulus and task features that seem critical for driving.

In another study by Ball et al. (1993), they used a large sample of older drivers to test this model, assessing various aspects of visual information processing including health status of the visual system, visual sensory function, visual attentional skills, and cognitive skills. The goal of the study was also to test a model designed to predict crash frequency in older drivers on the basis of visual and cognitive measures (Owsley et al., 1991).

The model as formulated postulates that eye health, central vision and peripheral vision have only indirect effects on crash frequency but direct effects on visual attention (UFOV). It further asserts that mental status has a direct effect on crash frequency, as well as an indirect effect on crash frequency mediated through UFOV.

The main role of central and peripheral vision in the model is their significant direct effect on the size of the UFOV; together central and peripheral vision accounted for 30% of the UFOV variance. Not surprisingly, visual attentional skills crucially depend on the integrity of information entering through the visual sensory channel. These results thus supported the hypothesis that UFOV is a mediating variable between crash frequency on the one hand, and eye health, visual function and mental status on the other. The analysis indicated that UFOV reduction is substantially better than chronological age at differentiating drivers who are at risk for crashes from those who are not.

There are several types of mechanisms that could potentially underlie a restriction in the size of the useful field of view in older adults. Earlier research has demonstrated that many older adults have deficits in selective attention and divided attention (Parasuraman and Nestor, 1991; Plude and
Hoyer, 1988), as well as a slowing in the rate of visual information processing (Hoyer and Plude, 1982). These types of deficits could contribute to a narrowing of the ‘perceptual window’.

Another potential cause of useful field of view constrictions is visual sensory impairment, such as severe loss in central and/or peripheral vision - an observer cannot attend to a visual event that is not adequately registered. Visual sensory and cognitive deficits in older adults can occur separately or together. It is important that because the useful field of view test relies on both visual sensory and cognitive skills, it provides a more global measure of visual functional status than either sensory or cognitive tests alone, thus improving its sensitivity and specificity in identifying older drivers at risk for crashes.

A.5.9 Driving experience and visual attention in car following when looking at in-car targets

It was shown that driving experience is indeed related to lane keeping when drivers have to perform a foveal in-car task continuously and rely on peripheral vision only in keeping the car within lane boundaries. A marked change appears in ability to do this successfully between the first 1500 and 50,000 km of driving (Summala et al., 1996). This is presumably connected with learning to use visual cues such as lateral speed, and kinesthetic and proprioceptive cues more effectively, but the actual perceptual mechanisms are still unknown (Riemersma, 1987; Smiley et al., 1980; Summala et al., 1996).

Such learning helps drivers to perform essential in-car tasks while keeping safely in a lane. However, drivers adapt their behaviour to changes in their skills, and improved lane-keeping performance may make them attend increasingly to non-traffic targets such as mobile telephones, other in-car accessories and sight seeing while driving (Summala, 1994).

Attending to in-car tasks may continue to be detrimental to detecting the braking of the car ahead for the experienced driver, however. Results show a substantial delay in brake reaction times when a driver is looking away from the lead car which brakes, whether he/she is still a novice or more experienced driver. In daylight, brake lights help little or not at all in detecting the lead car’s deceleration, if the following driver is looking at his/her speedometer or another target inside the car farther away from the lead car.

For a foveal in-car task in a given position, the retinal eccentricity of the car ahead (distance away from the central high-acuity vision) considerably exceeds that of the edge lines of the road, visible just above the dashboard, and other cues that a driver may use for lane keeping. It is assumed that
the looming retinal image of the car ahead is the major cue for detection of its deceleration (relative speed). The absolute perceptual threshold for detection of the relative speed in the fovea, in terms of angular speed of the approaching object, is ca. 0.002-0.003 rad/s (Hoffmann and Mortimer, 1996) which corresponds to the foveal thresholds found in these data at the 60 m initial distance (0.002 rad/s). It is presumed that this model works in peripheral vision as well (Summala et al., 1998). It is also suggested that the detection of deceleration of the car ahead when looking away from it is only dependent on the properties of the visual system. Although the use of peripheral vision can be practised (Johnson and Leibowitz, 1975), normal driving does not provide relevant practice for learning this task.

Practice in lane keeping also exceeds practice in detecting the braking of the car ahead, especially considering the use of peripheral vision, as the former is continuous and always functional in driving while the latter only occurs in car following. For such reasons distance-keeping with peripheral vision may not show much improvement with driving experience, as distinct from lane keeping. Drivers may not even have sufficient resolution in peripheral vision for detecting brake light onset of the car ahead (among other light flashes) while looking at the radio and other accessories on the mid-console.

Results support the hypothesis that detecting closing-headway situations with peripheral vision does not improve with driving experience as lane keeping does (Summala et al., 1996), at least within the range of 0 to ca. 100,000 km of lifetime mileage. It is presumed that no further improvement in peripheral detection will occur after this amount of mileage, which represents 1000-2000 hours of practice in normal driving, among ordinary private drivers at least. However, perceptual learning may occur in detecting deceleration of the car ahead in the foveal and parafoveal region, much practised in normal driving, resulting in lowered perception thresholds.

The fact that one in-car task causes two driving subtasks to deteriorate differently, considering visual performance capabilities, is inclined to mislead a driver with accumulating experience. It should be noted that improved lane-keeping capability is presumably one of the mechanisms which contribute to the subjective feeling of control.

A.5.10 Safety implications for using mobile (cellular) telephones while driving

A phone poll of 670 drivers commissioned by the Central Organisation for Traffic Safety in Finland in May 1997 (Liikenneturva, 1997) found that, indeed, 38% of the drivers had a mobile phone in
their car, with 24% using it daily whilst driving and 14% only using it infrequently whilst driving. Of these drivers, 42% felt they had increased their risk of having crash at some time while using a phone in the car, with 25% reporting a decrease in their attention to the road and other traffic whilst on the phone. Not too surprising was that 57% of the drivers felt that mobile phones in cars increase traffic safety, because they gave the ability to call home rather than speeding when late, call emergency services to crash scenes, and report dangerous road or traffic conditions to authorities.

Apart from the distracting task of dialling phone numbers and holding the phone when in use, there is also the distraction of conversation while driving (Lamble et al., 1999). Conversing on a mobile phone is unlike conversing with a passenger. Passengers normally have the opportunity to perceive the road situation and can vary the demands or timing of their conversation. Of course they may also distract the driver, as suggested by crash data from Doherty et al. (1998) but in their data the risk due to passengers is only increased in young drivers of age 16-19, possibly due to social interactions with peers. In the case of adult drivers the presence of a passenger even tends to reduce risk. When conversing on a mobile phone the person outside the vehicle does not have the opportunity to perceive what is happening on the road ahead. It is therefore possible that they could put the driver at higher risk than passengers do by demanding his/her attention when it is needed to cope with a critical situation in the traffic flow (Piersma, 1993).

Early experimental work has shown that perceptual and decision-making tasks are impaired when a driver has to divide attention between the road and a car phone (Brown et al., 1969). On-road studies have shown that hands-free phones cause less interference to the driver’s handling of the vehicle than hand held phones, however, hands-free phones still impair some aspects of driving performance (Brookhuis et al., 1991). Specifically, when drivers attempted to maintain a constant headway to a vehicle ahead and were engaged in a mobile phone conversation, their reactions to headway changes were somewhat delayed (Brookhuis et al., 1991).

A.5.11 Driving experience and time-sharing during in-car tasks on roads of different width

The results of a study by Wikman et al. (1998) support the hypothesis that driving experience, through more automated driving routines and processes of perception, means a more adequate allocation of visual attention while driving. This shows as a smaller variance in the length of glances at the secondary task (in-car task) among experienced drivers than inexperienced ones.
The way novices allocated their attention, with a large variance in glance lengths at the in-car tasks, has a practical consequence. The results showed that increased lengths of glances at the in-car tasks caused increased lateral deviation from the driving direction. No glances longer than 3 sec were taken by experienced drivers but such glances were found in 29% of the novices. The over-long glances of novices are therefore a potential safety risk, which is accentuated by the fact that the inexperienced, on account of less driving practice, are the least suited to look away from the road for long periods. Accordingly, it was shown that the novices were prone to extreme lateral displacements, increasing the risk of running off the road. Too many short glances, at least if they are not long enough to acquire essential visual information needed to perform an in-car task effectively, may not be desirable either, as they distract attention from the road unnecessarily. On the other hand, safe, short, orienting glances may contribute to task performance if they serve in planning the next glances.

The glances at the telephone (as an in-car task) were shorter on the highway than on the motorway. This indicates that drivers adapt their time-sharing strategy to the shorter time margins available on the highway by allocating shorter intervals of their attention to the in-car task. The percentage time during which the drivers looked at the in-car task also increased with the longer time margins on the motorway. Hada (1994) obtained similar results in an experiment where drivers were told to look at a display in the car for as long and often as seemed safe. The experienced drivers used significantly more glances at in-car tasks on the highway, which can be interpreted as a sign of better adaptation to the time margins among these drivers.
A.6 Dimensions of aberrant driver behaviour

A.6.1 Summary

Well-addressed topics on the subject of aberrant behaviour in the driving task relate to the distinction between errors and violations, based on the involvement of different failure mechanisms of cognitive competence together with motivational factors. Lapses and mistakes, as two psychologically distinct varieties of human error, are restricted to failures in individual information processing, while violations can be described as deviations within a regulated social context. Error recognition, correction, and signalling as stages on a continuum between automaticity and conscious control are employed to understand the nature of potential or inhibited errors, close shaves (sub-optimal or risky responses), and overt errors with regard to age and ability of people in demanding environments. To fully understand errors, not just actions or lapses, which have detectable or serious consequences, should be studied. Other work regarding driver error includes the importance of systematic and variable error on adaptation, and the important influence of conation, affect, hopes and fears as remote causes or favouring conditions preceding error. Research also touched upon the impacts of driver fatigue, alertness and so-called ‘cognitive tunnel vision’ on highway safety. Finally, questions surfaced about the validity of existing models to cover both cognitive and conative factors of (error in) human performance.

A.6.2 Errors and violations: dimensions of aberrant driver behaviour

There is now reasonable agreement (Reason et al., 1990b) that errors – defined as the failure of planned actions to achieve their intended consequences – can involve two psychologically distinct kinds of straying: the unwitting deviation of action from intention (slips and lapses), and the departure of planned actions from some satisfactory path towards a desired goal (mistakes). An inappropriate intention is a mistake and mistakes can be further subdivided, according to the skill-rule-knowledge taxonomy of Rasmussen (1980), into rule-based mistakes and knowledge-based mistakes. Slips and lapses concern the skill-based level of Rasmussen’s classification. Slips are potentially observable while lapses are more covert memory failures.

It is clear, however, that this error classification, restricted as it is to individual information processing, offers only a partial account of the possible varieties of aberrant behaviour. What is missing is a further level of analysis acknowledging that, for the most part, humans do not plan and execute their actions in isolation, but within a regulated social context. While errors may be
defined in relation to the cognitive processes of the individual, violations can only be described with regard to a social context in which behaviour is governed by operating procedures, codes of practice, rules, norms and the like.

Violations can be defined as deliberate (though not necessarily reprehensible) deviations from those practices believed necessary to maintain the safe operation of a potentially hazardous system. Violations need not always be formally prohibited but may also involve deviations from unofficial local norms of what is deemed safe or acceptable. The conceptual boundaries between errors and violations are not clear-cut and both errors and violations can be present within the same action sequence.

The finding that the subjects who report the most violations also tend to rate themselves as particularly skilful drivers suggests that these subjects believe that a good driver is someone who can bend the rules. Drivers who violate may see themselves as skilful enough to take risks (or perhaps, to behave in ways which would be risky only for less skilful drivers). Alternatively, drivers who violate may come to think of themselves as good drivers because they get away with what they do.

If it can be said that errors and lapses on the roads involve failures of cognitive competence (for example, failures of the effective, deployment of attention as a result of inadequate training, bad habits, or the distracting effects of mood), and that violations on the roads involve motivational factors, for example, trading the risk of a violation against the benefits of rapid progress, driving in accordance with a valued social image, or of indulging in self-expressive driving.

Slips and lapses are errors that occur at the skill-based level and concern behaviours that are performed more or less automatically. Automatic behaviour of drivers is overlearned and based on experience acquired over long periods of time, it is stereotypical and difficult to change. Improvement of automatic behaviour has been found to continue even after years of practice (Lewin, 1982). Therefore, at least some of the slips and lapses could be expected to develop as a result of increasing experience of the driver. For three-factor solutions (errors, lapses and violations) there have been no reports about an increase of aberrant behaviour with age or experience. Reason et al. (1990b) could only report a slight increase of lapses with age for older women but not for males. Parker et al. (1995) characterised lapses as the result of lack of concentration or attention, but there was no increase of lapses with age reported. Nor did Blockley and Hartley (1995) report any increase of errors or lapses with age. In contrast to the results of
previous studies, the four-factor solution presented by Åberg and Rimmó (1998) suggests that there are some slips and lapses that increase with age and experience of the drivers. The observation of a factor of inattention errors could be seen as an indication that driving failures to some extent might be caused by automatic behaviour. The number of inattention errors increased with number of years of holding a license and there are no differences between the sexes. Finally, females and drivers with low annual mileage reported more errors due to inexperience than other drivers did.

A.6.3 A model and proposed analysis technique for errors in driving tasks

The introduction of information-processing models of human behaviour represented a step forward in understanding how and why people fail to cope adequately with danger in the environment around them (Surry, 1969; Hale and Hale, 1970; Michon, 1985). However, such models have in the past tended to present a rather static and one-dimensional picture of behaviour. The division of information processing into the stages of input, processing and output, with specific factors feeding in at each stage, has tended to give the impression that the stages are carried out consciously and are therefore potentially equally open to direct influence. The models have not incorporated the long-established ideas that much behaviour comes with practice to be largely automatic, and is carried out with little involvement of either decision-making or conscious attention (Miller et al., 1960). In the process of learning and practising a routine task, plans or programmes are built up consisting of a number of closely coupled steps, which are then deployed as a whole in response to specific signals. Such plans take on a life of their own, and are sometimes difficult to interrupt or divert once set in motion. The plans are also difficult to modify once thoroughly learned, and may have to be demolished and relearned from scratch rather than adapted (Hale et al., 1988).

Rasmussen (1980) pointed to three different levels of functioning, called skill-based, rule-based and knowledge-based, based respectively upon the carrying out of highly practised routines, the choice of appropriate courses of action, and the development of new ways of coping with problems.

Reason (1976) conceives of the three levels as somewhat arbitrary points on a continuum of attentional control of behaviour. At the knowledge-based level, the full resources of attention are focused upon the chosen problem, because of its novelty or importance. Conscious processes such as hazard seeking, weighing of risk and other potential gains and losses, and a search for solutions take place at this level. Errors and accidents at this level are characterised by such factors as ignorance of risks, conscious preference for speed over safety, or preoccupation with one problem while another gets out of hand.
At the rule-based level, attention is allocated to the choice between already existing programmes, which are matched to the perceived needs of the situation. Errors are therefore characteristically diagnostic problems, where the individual is misled, either by external information or by preconceived expectations, into embarking upon an inappropriate course of action. These errors can be compounded by the failure to monitor the subsequent situation adequately enough to detect that the programme is leading towards, rather than away from the danger. Such incorrect choice is made more likely if distraction or preoccupation limits the amount of attention devoted to the process, in which case very rough approximations to expected situations may be accepted as triggers to set programmes in motion.

The skill-based level of behaviour includes those actions which are fully delegated to the automatic control mechanisms built into the learned programmes. Miller et al. (1960) conceptualised these checks as TOTE (Test-Operate-Test-Exit) loops, comparing desired outcome with actual state and then performing some action until the test showed an acceptably small discrepancy. No conscious attention is allocated to such monitoring, and the action only resurfaces into consciousness when major discrepancies are found by the built-in monitoring. Indeed, premature interference by the attention mechanisms can actually disrupt the smooth functioning of the programme and cause it to lose its place or rhythm. The steps follow each other automatically, governed by some mechanism of transitional probability. Large numbers of these programmes are learned during the course of experience at any activity, and practice makes them highly efficient routines. Where the same objectives have to be achieved in differing circumstances, or different objectives in the same circumstances, very similar programmes with shared steps will be built up. Reason points to a number of characteristic types of error in such families of tasks:

a) Cross-talk between similar programmes leading to branching into the wrong one.

b) Capture of behaviour by a programme triggered either by external circumstances or by internal preoccupations.

c) Losing the place in a sequence through distraction and either missing or repeating a step, or a monitoring loop.

Within the learned programmes, motivational factors have only a very small role to play in the behaviour itself, though they will have had a role in the process of learning the programmes. The monitoring schedule applied to the behaviour, and hence the chance that major discrepancies will be picked up will be under more conscious control, but the dynamics of the situation will play the major part in determining whether such monitoring can be effective (Hale et al., 1988).
A.6.4 A basic driver error: late detection

Rumar (1990) considered detection errors: that is, failure by a road user to detect another road user in time to be able to avoid him/her while successfully competing a planned course of action.

In road traffic, with a considerable physical and human inertia, it is obvious that failure to detect the other road user early enough is an important source of error. This conclusion is also supported by explanations of traffic accidents by people given in court. The most frequent explanations for such accidents are for example, “I saw him too late”, “Suddenly he was there”.

Failures to detect another road user early enough to avoid a collision may have several causes, but seem to belong to two main categories:

a) Cognitive errors, illustrated by a failure to look in the direction or for the specific type of the road user in question;

b) Perceptual errors, illustrated by the failure to detect another road user in peripheral vision or in situations of reduced ambient illumination.

The attention, the focus of interest, must often be directed voluntarily in a conscious, planned, and controlled way. According to the rules of the road, the critical vehicle often appears from the right (the right hand rule in most countries with right hand driving). The search patterns learned and used to detect oncoming vehicles on roads are not automatic and skill-based, but controlled and rule-based (Rasmussen, 1980). As such, the patterns are not as efficient and quick as natural skill-based patterns. They are more vulnerable to errors, and due to the speeds and the masses involved it is very difficult to compensate for an error once it begins to occur. Errors are thus both more probable and more difficult to overcome.

There are situations where the natural skill-based behaviour is in direct conflict with the rule or knowledge-based, officially-required behaviour. Such clear or contradictory traffic conditions, of course, increase the probability of an error occurring, for both natural and automatic as well as controlled and rule or knowledge-based detection errors.

Another detection error of cognitive aetiology occurs when a road user has a partly correct expectation. The driver has a functional model of the traffic which is appropriate in terms of when and where to look, but inappropriate in terms of what to look for. That is, a model which is correct
in spatial and temporal relationships but wrong in terms of target selection. The road accident history is full of such examples. Road users usually look in the appropriate direction at the appropriate moment, but they also usually look for cars. This is a very natural behaviour since cars are often the largest, heaviest, fastest and consequently most dangerous moving object visible. Unfortunately, after doing this and seeing no car within a critical distance, they complete the action they had planned and hit or are hit by a cyclist, a motorcyclist or even by a pedestrian. This problem of ‘looking but not seeing’ in road traffic has not been investigated except in the case of road signs (for example, Johansson and Rumar, 1966). This is a cognitive detection error which probably has a considerable motivational component.

A third detection error of cognitive aetiology may be caused by internal distraction. People are preoccupied by some personal problem to such an extent that their normal information search patterns are destroyed or they are looking without seeing.

Cognitive detection errors of the type discussed above are a result of a discrepancy between the road user model of the traffic environment and the real situation. This may lead to erroneous expectancies, or lack of expectation, which in turn may contribute to failure to detect the other road users in time to avoid a collision.

It is argued that perceptual detection errors have an evolutionary history. Stimulus patterns which were essential for detection during our evolutionary history and which determined the development of our visual perceptual systems are not available in many contemporary road traffic situations. Peripheral sensitivity to movement is ineffective since the contours of our present ‘enemies’, i.e. cars, have no perceptible internal motion. They roll closer without any associated higher frequency movements, such as the leg movements of an animal predator. The contrast sensitivity of the eye is so impaired at lower levels of ambient illumination that other road users without lights are thus not detected early enough in night traffic.

By improving the conspicuity of road users, it compensate for both types of road-traffic detection errors - cognitive as well as perceptual. However, it is risky to focus all efforts at countermeasures on improving detection by enhancement of stimuli conspicuity. By continuing to enhance conspicuity, one arrives at a point where the more powerful cues being provided begin to compete with, and to mask each other.
A.6.5 Error recognition, correction and signalling

It has become a convenient convention in applied psychology to discuss errors as binary events, which either do, or do not occur. In real-life, the definition of a lapse is much fuzzier. A tired driver may abruptly become aware of dangerous drift across a road – but with no immediate consequence or alarm if there is no other traffic. Thus, the likelihood that a lapse of memory, or attention will produce a memorable error is a function of how forgiving is the environment in which it occurs. Errors can be defined as overt completions of inappropriate actions which may have undesirable consequences, or which may increase the probability of undesirable consequences (Rabbitt, 1990). However, in offering this definition, it must also be recognised that people may often internally programme inappropriate responses which they nevertheless manage to inhibit before they complete them. They may also plan, and successfully complete, responses which do indeed achieve the results that they intended – but in a sub-optimal way. In an unforgiving environment, a person who becomes aware that he/she is increasingly often suppressing potential errors, or increasingly often making sub-optimal or risky responses, will take this as an important danger signal of a slippage in efficiency which, if not checked or compensated, may lead to a serious error. Thus in more demanding environments, such as those more usually inhabited by younger people and by individuals with higher IQs, it is vital to continually monitor one's performance to detect and analyse ‘close shaves’ and small deviations from perfect performance as well as overt errors. The retired elderly, and individuals with lower IQs, more usually inhabit lenient environments in which danger signals of drift from perfect performance can be ignored without hazard – until a very improbable and potentially dangerous event abruptly occurs, leaving no chance of recovery.

Error recognition, error correction, and error signalling appear to represent a hierarchy of progressively less ‘automatic’ – and increasingly slow and decreasingly reliable – process. The automaticity of error correction responses is evidenced by the fact that they apparently cannot reliably be consciously suppressed. Even when subjects are specifically instructed to ignore all of the errors that they make, they frequently produce pairs of sequent errors because they fail to suppress spontaneous correcting responses, i.e. they make the response they should have made on the last trial without reference to any new signal which may since have appeared. A taxonomy of error detection, error correction and error-signalling responses as stages on a continuum between automaticity and conscious control is interesting because it focuses attention on qualitative characteristics of error detection and error signalling responses other than their obvious quantitative differences in speed and accuracy. It also encourages interesting new predictions.
detection responses (post-error slowing) and error correction responses (making the response one
should have made) are highly automatic processes, they will also be relatively invulnerable to
interference from secondary tasks, relatively unaffected by gross changes in information processing
load and, most importantly for the present discussion, relatively inaccessible to consciousness and
relatively poorly remembered. In contrast more consciously controlled error signalling responses
will be relatively vulnerable to secondary task interference, relatively affected by main task
information processing load and, allowing for the fact that they are most rarely produced than error
signalling responses, they will be relatively well remembered.

There is no evidence that people’s ability to automatically register their error changes as they grow
old. In all error signalling conditions, people of all ages and IQ ranges seem equally able to
recognise nearly all the errors that they make at some automatic level of information processing.
However, after they have completed an experimental run they appear to remember more of the
errors which they have signalled by making overt, controlled error-signalling response than of the
errors that they have only automatically detected or corrected.

A final point strongly emphasised by these choice reaction time (CRT) experiments, is the
distinction between the relative memorability, for people of all ages, of acknowledged as against
unacknowledged errors. Koriat and Hasida (1988) make a very similar point in their illuminating
analysis of factors which make some slips of action more memorable than others. They suggest
that uncertainty of recollection as to whether an habitual, and so highly automatic, action (such as
locking a door) has, or has not, actually been completed may be more easily resolved if it happened
to be accompanied by some other, irrelevant but unexpected event (such as accidentally kicking
over a milk bottle on the doorstep) than if it proceeded smoothly, and without any extrinsic marker
to distinguish it from hundreds of other almost identical acts. To this interesting speculation that
unexpected diversions from otherwise unmemorable instantiations of highly automatic, stereotyped,
actions may act as memory markers, Rabbitt add evidence that when errors made during a highly-
automatised and repetitive task are immediately and overtly acknowledged they will be better
remembered, and used as evidence for self-assessments of reliability than if they are not marked by
overt, unique controlled acts.

For obvious reasons analytic frameworks for quantitative estimates of human errors as contributory
causes of disasters and accidents, or analyses of slips of action (Reason, 1979) have been based on
the production of overt, and so countable, ‘actions not as planned’. A striking limitation of this
framework of description is that actions are not classified as errors unless they have some undesired
outcome. However, field studies often reinforce the point that complicated disasters usually do not have unique causes, but rather occur because of complex and rare concatenations of factors. A lapse by a human operator which might, in another context, have been entirely inconsequential is often only one of many factors in this rich pattern of causation. One cannot learn very much about errors if one studies only lapses which have serious, or at least detectable consequences. Operators who begin to make increased numbers of lapses which have no immediate consequences and which may, individually, be retrievable by corrective action, nevertheless thereby steadily increase the risk that, sooner or later, a coincidence of one of these slips with a concatenation of other, improbable, events will escalate into an irretrievable blunder.

Laboratory studies of self-evaluation in everyday life begin to show marked individual differences in the stringency, sensitivity and vigilance with which people monitor and control their performance of simple tasks. Obviously, in real life, the perception of the need stringently to monitor one’s own performance, and so the effort put into it, will vary with the situation. A person pouring acid from a thermos flask will probably notice, strive to correct and control – and subsequently remember slight hand-tremors far better than will a person pouring tea. However, these studies suggest the point that even in situations which make identical demands on self-monitoring there are likely to be individual differences in the criteria adopted for ‘riskiness’ and in the sensitivity with which deviations from these criteria can be detected.

Rabbitt (1990) also warns about the many differences between laboratory memory tasks and the demands of everyday life. Laboratory memory tasks are experienced under conditions of minimal distraction and stress. In everyday life, competing activities and distraction are the norm. In everyday life, people depend on overlearned domain-specific memory organisation structures (for example, Ericsson and Polson, 1988), and on practiced cognitive skills which they cannot develop or use during their brief experiences of novel laboratory tasks. Unlike tasks in everyday life, laboratory experiments are typically very brief and so do not require sustained attention. Most importantly, unlike everyday memory tasks, most laboratory tasks do not test the efficiency with which people can switch between competing activities or remember and use plans to control complex sequences of actions.
A.6.6 Variable errors setting a limit to adaptation

What is it about human error that causes unexpected events which are called accidents, and why do humans not learn to eliminate errors? Brehmer's (1990) argument is that one simply cannot do so because at least part of what is called human error is due to inherent variability.

It seems that there could be two preconditions for errors leading to accidents. One is that people somehow misperceive the situation, and therefore choose an incorrect action. For example, a driver may seriously underestimate the speed of an oncoming car, and decide to overtake when he/she should not have done so. Alternatively, one may fail to carry out an action as planned. For example, it may take longer to stop the car than one thought. This is the basic distinction between perceptual errors and errors of execution. The former lead us to believe that the situation is different from what it actually is, while the latter lead us to believe that we are able to do something that we cannot do.

Both perceptual and execution errors may be of two kinds: systematic, or constant errors, and variable errors. The former kind of error is defined as the systematic difference between some target value and the actual response. In practice, this is defined as the difference between the mean of the responses and the target value. Variable error, on the other hand, refers to the variance of distribution of the responses. The latter error is, of course, independent of the former, and it is possible to have variable error without constant error.

The current focus in research on human error is upon a number of well-defined constant error tendencies. The most popular explanation for these systematic tendencies is that humans have limited information-processing capacity, i.e. that people have to act in an environment that is too complex.

Perhaps the neglect of variable error in modern psychology, according to Brehmer, comes from the fact that psychological experiments are seldom set-up to examine variable error (rather than examining what a person does repeatedly in the same situation), presumably because the between-subjects variability thus estimated can be used as an error term in the statistical analyses. Any replication within subjects would not be of any obvious use in such tests, and it is therefore not examined in the experiment. Moreover, even though people exhibit variable error, only part of the distribution is actually noticed.
Systematic error does not introduce any limits to adaptation; it is possible to compensate for constant error, and humans are quite adept at doing so (Chapanis et al., 1949). It seems that constant error cannot really be the cause of accidents, except, perhaps, for inexperienced operators.

Variable error is different. People are generally unaware of the fact that their performance varies and, even if they are made aware of it, they cannot predict which state they will be in at a given moment in time. Variable error exists everywhere in the human system: on the input side in perception, in the mediating systems, for example, in judgment, and in the output side in action, for example, in reaction times.

The results from experiments support the hypothesis that subjects will treat uncertainty due to variable error in the perceptual system in the same manner as they treat uncertainty located in the physical system. This supports the general hypothesis that variable error in the perceptual system turns the deterministic physical driving environment into a probabilistic one. Variable error, therefore, introduces a limit to the extent to which a person may adapt.

A.6.7 Prospects for technological countermeasures against driver fatigue

Adverse effects of fatigue on self-assessments have been reported for many years in the psychological literature. Bartlett's (1948) studies of pilot fatigue among World War II bomber crews showed how skilled performance disintegrated over time, with tired individuals reacting to isolated components of their task, rather than to the task as an integrated whole. Since then, many laboratory studies on watch keeping and vigilance have confirmed these effects on attentional distribution as repetitive work is prolonged. The onset of so-called tunnel vision will impair the ability of individuals to process feedback about the effects of their actions on the task, making it extremely difficult for them to detect deteriorations in skilled performance, even though they may be well aware of feeling tired.

On-road studies (for example, Brown, 1970) have also shown that choice behaviour becomes more risky as a consequence of fatigue. Hence, even if drivers know they are drowsy, they are likely to take the easy option of finishing their journey, rather than face the wrath of employers and customers, and confronting the problem of finding accommodation with somewhere to park their vehicle.
In addition, the problem exists that, although fatigues drivers usually seem to know when they are becoming excessively sleepy, this knowledge does not necessarily help them judge how unalert they actually are. Iizuka et al. (1985) obtained neurological and physiological measures of alertness from drivers who also provided subjective assessments of drowsiness. They found that feelings of drowsiness were a completely unreliable indication of a driver’s true attentional state at very low levels of objective alertness. This finding has, to some extent, been confirmed by Horne and Reyner’s (1995) simulator study of driver fatigue, showing that, in the latter stages of declining alertness, drowsy drivers may not realise just how close they are to actually falling asleep at the wheel. As was pointed out by Brown (1995), this means that tired individuals may remain unaware of the risk they are taking in continuing to drive until they begin micro-sleeping and become conscious of lane drifting, or near collisions; always assuming they survive the consequences of their initial micro-sleep.

For reasons such as these, it may be concluded that very fatigued drivers are not reliable decision-takers, particularly about the extent or consequences of their own impaired skills (Brown, 1997).

A number of studies (for example, Richardson, 1995) have explored the possibility of employing systems which learn to recognise a driver’s normal pattern of responding and then identify significant departures from it. There is no problem with this approach, using a neural net type of pattern-learning procedure. However, driving behaviour is a changing function of task and environmental demands and it therefore seems likely that multiple inputs of behaviourally-related information from these sources will be necessary, if any technological system designed for driver fatigue detection over long journeys is to be reliable and if it is to have any credibility among potential buyers and users. For example, the system may find it useful to know something about road and traffic conditions, about time into journey, and about time since sleep, so that measurement criteria could be adjusted to take account of these mediating influences on driver performance, allowing the reliability of the system’s output to be maximised whatever the driving conditions (Brown, 1997). Some of this information is, of course, already routinely available from the tachograph fitted to many commercial vehicles.

A.6.8 Designing for transportation safety in the light of perception, attention and mental models

It is customary to treat the detection of a signal as a problem in maximising detections while trying to minimise, or at least keep to some specified level, false alarms. However, reflecting on the nature of false alarms, normally thought of as errors, brings out an interesting point, relevant to
Reason's (1990a) treatment of mistakes. A mistake is by definition a choice of an inappropriate intention. Is a false alarm, or indeed a miss, an error at all? Moray (1990) considers the basic paradigm. The observer knows that sooner or later an event will happen which may contain a signal that he/she must detect. Note that the event may be one determined by the environment. It may be a light or sound generated by an experimenter, or it may be the appearance of a traffic signal of a particular colour and shape. On the other hand, the event may be generated by the observer directing his/her eyes to a particular part of the environment. The observer knows enough about the task to know what the signal probability is, and the payoff structure of the situation, in terms of the expected value or utility of hits, misses, false positives and true negatives. If one places the observer on a particular ROC (receiver operating characteristic) curve, and specify expectations and payoffs, then a false positive is not an error. Rather, it is a tactical decision about what to report. "I have been told that I should not miss more than 5% of the signals, and given what I know about the situation, I shall therefore say 'yes' when I am uncertain in about 5% of the cases'. To say 'yes' when uncertain, given that choice of tactic, is not an error. It is not a mistake, and it is certainly not a slip. It is a correct rational judgment. This is true even of judgments about such things as the presence or absence of vehicles or pedestrians on a road, adequate space to overtake, or indeed any perceptual judgment.

This example draws attention to the role of tactics in information processing, and emphasises their importance in understanding mistakes. Consider the following results which were recently obtained in an experiment on process control (Moray and Rotenberg, 1989). The task was a simulation of a simple thermal hydraulic system based on Crossman's 'water bath' task (Crossman and Cooke, 1974). There were four tanks, each with an independent control of input and output, and each with its own heater. The task was to bring each tank up to set points for temperature, level of fluid and outflow. The set points for each tank were different. Operators performed the tasks for several hours until they had reached a stable level of performance, at which time they could bring the tanks to set point in open loop mode from cold start-up. However, there were small random variations in the state variables, and these required the operators to trim the control variables from time to time. Without warning, a fault developed in one tank. The operators began to try to compensate for the disturbance, and about 75 sec later, a fault developed in another tank, which also required immediate attention. The operators were completely absorbed by the first faulty subsystem. The proportion of time that they spent viewing the other subsystem was drastically reduced for several minutes, and as a result in all cases the second fault went unnoticed, or at least untreated, for several minutes, during which time its state variables typically diverged from their set points to a very large degree. Interestingly, the operators sometimes looked at the
subsystem with the second fault, but did nothing about it. Either they did not notice its anomalous state, or they refused to time-share fault management, deciding instead to complete the management of the first fault before starting on the second. If the latter interpretation is correct, it has serious implications for risk management.

This kind of behaviour is typical of what is called ‘cognitive tunnel vision’ (CTV) (Moray, 1981), ‘cognitive lock-up’ (Sheridan, 1987), or ‘cognitive bias’ (Kahneman et al., 1982). It appears at first sight to be erroneous behaviour, since the operators pay attention only to a small subsystem of the entire plant, and fail to distribute their attention over the entire system. Examples of similar behaviour have been reported in a number of cases of transportation and other industrial accidents. However, in many cases where one has been inclined to see CTV as a major cognitive error, such a judgment may be incorrect.

Given that the observer is confronted with an anomaly in one subsystem, this means that it is far more likely that the cause of the anomaly and its cure will be within the set of tightly coupled components of the subsystem within which the fault is signalled, rather than in some distant and loosely coupled subsystem. When faced by the limited attentional resources available to them it makes good tactical sense for operators to confine their attention to a single subsystem, at least until they become certain that the fault is not actually in that part of the system. There is a good case for saying that CTV is actually a correct adaptive strategy rather than an error. At first sight such behaviour, based as it is on a prolonged experience of the structure of the system under normal operation, is a case of what Reason calls ‘Strong but Wrong’ habitual behaviour. Viewed from the point of limited resource allocation however, it is perhaps more a case of ‘Rule but Cool’. This point of view is supported by the work of Klayman and Ha (1987).

**A.6.9 Current and future perspectives on error analyses and traffic behaviour**

One can look at error and accidents from the perspective of risk taking. Although the possible occurrence of an accident is considered, it is outweighed by other factors such as economic advantage or minimisation of effort. Having followed the many discussions around this concept Lourens (1990) argues that there are more and better reasons to abandon the concept altogether than to let it play its confusing role in a model of human action. Trying to define the concept of risk has proven to be an almost useless exercise (Doderlein, 1987).
Explanations of performance errors in terms of mechanisms of risk, decision-making, or problem solving will very often be incomplete and insufficient for the development of truly effective measures. This is in agreement with Kjellén (1987), who stated that there is no single theory or model which predicts the occurrence of human errors and their effects, that may serve as a sole basis for data collection and analysis. A comprehensive approach involves different theories and models which consider: (i) the effects of physiological and motivational factors of the working environment on the probability of human errors, and (ii) the human processing of information on deviations from a goal and the prerequisites for the identification and correction of these deviations by the human operator.

Prominent cognitive psychologists recognised this some time ago. Neisser (1963) noticed the conspicuous absence in cognitive models and computer simulations of the facts, that (i) human thinking always takes place in, and contributes to, a cumulative process of growth and development, (ii) human thinking begins in an intimate association with emotions and feelings which is never entirely lost, and (iii) almost all human activity, including thinking, serves not one but a multiplicity of motives at the same time.

Selecting or choosing is carried out through weighing the desirability and feasibility of wishes (values and expectancies). In general, it can safely be assumed that people are not concerned – as Hale and Glendon (1987) put it – to ‘squeeze the last drop of advantage out of a situation’, but merely to accept an outcome which is good enough or satisfying.

Hale and Glendon summarised some personality-related emotional influences on subsequent behaviour. They noted that:

a) Aggression or anger will make people less willing to put up with inconvenience and frustration from safety precautions;

b) Anxiety or fear result in more obsessive checking of action sequences, thereby interrupting the smooth flow of routine actions.

There are a number of researchers who focus their work on the output side of human behaviour and try to build what they call ‘action theories’. A key to the action theory perspective is an interest in data on the interplay between observable behaviour on the one hand and covert cognitive and emotional processes on the other.
Norman, (1980,1981) proposed his ATS (Activation-Trigger-Schema) theory of action, which claims that action sequences are controlled by sensorimotor knowledge structures, the so-called schemas. Most actions are carried out by subconscious mechanisms. At conscious level one prepares only the general selection of the act: that is one ‘wills’ an action, while the lower-level components complete the action to a large extent, without further need for conscious intervention, except at critical choice points. Schemas can be thought to function at different levels. There are high-level schemas (‘parent’ schemas) which cover a compound activity such as driving a car, and low-level schemas (‘child’ schemas) which cover single actions within as activity such as car manoeuvres. At any one time, numerous schemas may be active. Schemas however, only invoke actions when they have been triggered, and this requires satisfaction of trigger conditions plus a sufficiently high level of activation. This might explain why emergency reactions do not always come up in time: in acute danger, the relevant emergency schemas are triggered, but not activated soon enough. The person is caught by surprise.

The theories described so far provide relevant information to explain many cases of human error, if by explaining is meant giving a plausible indication of the psychological source of error. Many psychologists and practitioners would rightfully be perfectly happy with this level of explanation. But it is possible to go one level deeper – it always is – and to try to explain why these sources of error are there in the first place. Then, the problem of the origin of human error is addressed. Going back in history, Spearman (1928) had some very fundamental things to say about the origin of error. Spearman explained the origin of error by what he called the doctrine of ‘noegenesis’: the laws and processes whereby belief is derived from intuitive evidence.

For Spearman it is unquestionable so, that conation and affect play an important part. A very large share of mistakes that a person makes in life, are due to allowing beliefs to be swayed by hopes and fears. Nevertheless, the influence of conation and affect, however potent it may be in degree, is never really the proximate cause of error, but only a remote cause or a favouring condition.

According to Lourens, therefore, models of (error in) human performance should explicitly cover both cognitive and conative factors. Whether such a plan calls theoretically for an integration of existing models or for a new paradigm in research, was difficult for him to say. Probably a new metaphor is needed. The computer is fine as far as ‘human computation’ is concerned. To do justice to the whole of the human being, no suitable artefact is available for a metaphor, perhaps for the use of the human brain itself.
A.7 Driver Behaviour Modelling

A.7.1 Summary

More than a decade ago, researchers have noted the lack of progress in developing a comprehensive model of driving behaviour (Michon, 1985; Huguenin, 1988). The reasons which have been proposed for the stalemate back then (Ranney, 1994), seem to be still explanatory of the situation many years later. These include a preoccupation in the highway safety field with accidents and accident-causing behaviours. As a result, it has not been clear whether theories should explain everyday driving, or accident-causing behaviours, or both. Secondly, motivational and risk models, which emerged in the 1960s and 1970s as alternatives to skill-based models of driving, have failed to generate testable hypotheses necessary for developing a body of empirical findings. Importantly, some of these models have been criticised for failing to distinguish between the aggregate (or macroscopic) and individual levels of analysis (Michon, 1989). Much valuable effort has been spent on protracted debates on the validity of these theories. In addition, the cognitive revolution in psychology has failed to influence driving behaviour modelling (Michon, 1985), not to mention the equally important revolution in research on emotions recently (LeDoux, 1998). Ranney (1994) gives a comprehensive overview on the state of affairs in driver behaviour research up to 1994, and explores more recent avenues of research in driver behaviour modelling, such as the importance of visual search and attentional mechanisms, cognitive processing, automaticity and error-correcting behaviour.

Within the traffic engineering field, modelling the behaviour of individual drivers in computer simulation software offers engineers an alternative approach to macroscopic models for road infrastructure design and analysis. Advantages of this type of microscopic simulation include a better understanding of traffic patterns and underlying processes, together with the ability to predict traffic conditions which have not been observed. Driver behaviour is mostly based on information-processing models incorporating if-then rules, decision trees, deterministic car-following and lane-change models, and inadequate mechanisms for learning (Oza, 1998). The complication that driver actions, chosen when values are on either side of decision thresholds, need to be similar enough to prevent unstable oscillations, burdens these models even more. Furthermore, the transfer of new insights in driver behaviour modelling from the field of traffic psychology into the software of driver simulators for traffic engineering applications, seems to be absent. The development, testing, and safe implementation of electronic driving aids and transport automation necessitate and depend on humanlike driver behaviour of autonomous drivers in driving simulators, and accurate
modelling/prediction of individual driver behaviour in simulation software and on-board monitoring/warning systems.

A.7.2 A Review of driver models (Ranney, 1994)

Individual differences in accident causation

In the highway safety field, priority has generally been given to identifying risk factors through epidemiological studies of accident causation. The result has been an over reliance on accidents and accident-causing behaviours, and a failure to consider driving behaviour within the broader context of transportation for a particular purpose (for example, to get from home to work). Much experimental work has attempted to identify individual differences in basic capabilities that predict accident involvement, for the purpose of selecting drivers with above-average crash risk. This research has been referred to as the study of differential accident involvement (McKenna, 1982) or the individual differences or selection approach (Barrett et al., 1973). The emphasis has been on identifying relatively stable traits, as opposed to the more transient stated emphasised by motivational models of driving (Johnston and Perry, 1980). Accident proneness was the focus of much of the early work on driving behaviour, although this concept offered little in the way of explanatory potential. McKenna (1982) considered the prediction of accident involvement on the basis of psychological tests to be an improvement over statistical attempts to identify accident-prone individuals, because it offers the potential for a theoretical understanding of the psychological abilities and characteristics associated with the errors involved in crash causation.

Early attempts using simple visual attributes and simple reaction time as predictors found weak or no relationships with accidents, presumably because drivers can compensate for deficiencies in these abilities (Hills, 1980; Summala, 1988; McKenna, 1982).

Harano et al. (1975) combined biographical (age, sex, occupation), overt behaviours (prior violations, annual mileage) and psychological attributes (perceptual style, attitudes) in predicting previous accidents. Psychomotor and perceptual variables were considerably less important than biographical and exposure factors. Marital status, mileage, traffic conviction record, and socio-economic status were among the more significant predictors. Barrett et al. (1973) suggested that combining different categories of constructs could lead to conceptual confusion, due to the possibility of complex interactions between constructs at different levels of measurement. While the study by Harano et al. was clearly conducted without concern for the processes underlying the
identified predictive factors, the objective was to improve driver-screening procedures and not to provide an understanding of accident causation.

Perhaps the most important work relating to the identification of predictors of accident experience was done by Barrett et al. (1973), who presented a conceptual analysis based on accident-cause data supporting their conclusion that three categories of information-processing measures were relevant for predicting accidents. These included perceptual style, selective attention, and perceptual-motor reaction time. A later analysis by Panek et al. (1977), focusing on the effects of aging on driving, provided rationale for adding vision, vigilance, and decision making, thus broadening their model to cover driving behaviour rather than just accident causation.

The strongest and most consistent predictor, deriving from the work of Barrett et al. (1968, 1973), is selective attention. Based on the assumption that rapid switching of attention is required for complex task performance, such as driving and flying, numerous studies have found significant correlations between measures of selective attention and accident involvement (Kahneman et al., 1973; Mihal and Barret, 1976; Avolio et al., 1985). The majority of these studies have used a dichotic listening task (DLT) also referred to as the Auditory Selective Attention Task (ASAT), developed by Gopher and Kahneman (1971). The test requires subjects to respond to strings of letters presented simultaneously to each channel (ear). Each message has two parts, so that on some trials the instructions change between parts and require a rapid switch of attention from one channel to the other. The number of omissions, intrusions, and switching errors are recorded.

Most recently, Owsley et al. (1991) have found visual attention to be a significant predictor of retrospective accidents for older subjects. Their measure of attention, referred to as the useful field of view (UFOV), is a composite measure of preattentive (parallel) processing, incorporating speed of visual information processing, ability to ignore distractors (selective attention), and ability to divide attention. Their work was based on a hierarchical model of vision and information processing in driving, which included factors at the following levels: ophthalmologic (eye health), functional vision, preattentive (UFOV), and cognitive function (mental status). An important feature of this work is the use of disaggregated accident data. Specifically, strong associations were found between the UFOV and culpability in intersection accidents. In a subsequent larger-sample study, significant correlations were observed between UFOV and all types of accidents (Ball et al., 1993). The relatively high correlation values observed by Owsley et al. (1991) may reflect their use of selected samples, for which correlations are generally expected to be larger than would be found in the general population (Harano et al., 1975).
What do tests really measure?

The use of simple correlational methods without multifactoral structural models raises questions about the meaning of significant correlations (Kenny, 1979). In the absence of underlying theoretical models, post-hoc explanations have been proposed to explain the relationships between predictors and criterion measures. For example, it is often cited that “driver attention” is responsible for a large percentage of accidents (Zaidel et al., 1979). The finding, based on police officers’ judgments, is used to justify efforts to find correlations between laboratory measures of attention and accident measures (Parasuraman and Nestor, 1991). This analysis mixes scientific with everyday concepts of attention and serves to reduce the construct to the most basic common denominator used in everyday language, as in “paying attention”. Although this would appear to restrict researchers from taking advantage of theoretical distinctions between mechanisms of sustained, selective, and divided attention, in practice it has had the effect of allowing researchers to claim generality based on successes of different measures of attention. For example, the dichotic listening task (DLT) and the useful field of view (UFOV) are both tests of attention. Positive correlations between both tests and accident measures have been observed, however, the two tests evaluate different mechanisms. Both tasks involve specification of targets and the individual’s ability to distinguish targets from distracters. However, the UFOV uses very brief exposure durations and is thus concerned with preattentive or the earliest stage of visual processing in which attention is captured and directed to salient visual events. Deficits of information-processing speed, which are evident among older individuals, can contribute significantly to performance on this task (Ball et al., 1990). In the contrast to this bottom-up (data-driven) processing, the DLT emphasises the strategic control of selection and switching between channels, and thus reflects top-down (memory-driven) processing. Furthermore, it is questionable whether either test addresses the same mechanisms cited as inattention on police accident reports. Poor operational definitions of attentional mechanisms have been identified as a shortcoming of test-based research in the highway safety field by Sivak (1981).

Methodological considerations

Barrett et al. (1973) identified several potential problems with the “unusual” postdictive research design that appears to be unique to the field of accident analysis. These include: (i) restricted range of criterion and/or predictor variables, for example, as might be due to the death of the worst drivers before they can be tested, (ii) the potential effect on motivation or test performance of knowledge...
by the driver of having been placed in a special category due to accident involvement, and (iii) the questionable assumption that skills or attributes measured by the individual variables are highly reliable and do not change over time.

Rabbit (1981) criticised psychometric research for undue emphasis on the development of reliable predictors, without concern for the underlying psychological processes.

Accidents as a criterion

Potential problems with the use of accidents as a basis for evaluating individual drivers include their lack of stability, the lack of statistical power due to their infrequency, and the lack of reliability of state accident records. The stability of accident data has been considered in several studies. Burg (1970) recommended against trying to identify accident repeaters since the vast majority of accidents in one time period involve previously accident-free drivers. Arthur and Doverspike (1991) found different predictors of retrospective and prospective accidents, indicating a lack of stability of accident involvement.

Because accidents are relatively infrequent events, data from a number of years must be combined to obtain sufficient statistical power. However, Evans (1991) has demonstrated that the extremely low likelihood of accident involvement for ordinary drivers creates significant difficulties in detecting drivers with above-average crash likelihood, even when periods of up to seven years are used. Use of longer periods is not recommended due to the instability of accident records, mentioned above. This is another possible explanation for the relatively low correlations normally found in studies of retrospective accident prediction.

The reliability of accident records has been questioned in a study conducted by the Insurance Research Council (1991). Over 27,000 accidents from 40 states in the USA from 1990, serious enough to meet the respective state reporting requirements, were included in the sample. It was found that only 40% of these accidents appeared on official state records. When compared with an earlier study, it was found that the reporting percentage had dropped from 48% in 1983 to the 40% in 1990. It is questionable whether data collected for administrative purposes adhere to standards necessary for research.

Due to the shortcomings of accident data, it is not uncommon for researchers to recommend use of alternative criterion measures (Johnston and Perry, 1980; Sivak, 1981; Mckenna et al., 1986).
However, this is considerably easier said than done, based in part on the conclusion that such measures would necessarily require validation against accidents (Ball and Owsley, 1991).

Several researchers have proposed that alternative criterion measures be developed without the requirements of validation against accident measures. Sivak (1981) proposed that skills with high face validity to driving be selected and evaluated by examining the effects of transient human states (fatigue, alcohol intoxication, stress). Skills sensitive to the effects of transient factors would thus be considered most critical for driving. Following their difficulty interpreting correlations between different laboratory measures, McKenna et al. (1986) concluded that identifying causes of error would be a better research focus than trying to predict past accident. Similarly, based on positive correlations between observed errors, their level of danger, and accident incidence, Brown (1990) concluded that field-testing of hypotheses developed from theories of driver error is a more valid approach to highway safety than the reliance on pot-hoc subjective assessments of error contributions to accident statistics.

The available evidence clearly supports the use of alternative criterion measures, yet surrogate measures are not widely used. This reflects both the failure of existing theoretical models to provide such measures and the uncertainty associated with interpreting such measures if there is no connection to safety.

Functional models of driving behaviour

Michon (1985) distinguished between taxonomic and functional models of driving behaviour. Trait or test-based models, which underlie most of the individual differences and accident causation research, are taxonomic, and thus involve no dynamic relations among components. In contrast, functional models, which include motivational models and information-processing models, have greater potential for helping to understand complex tasks such as driving (Michon, 1985).

The demand for a comprehensive model of driver behaviour is not well satisfied by theories of traffic psychology (Huguenin, 1988). Several different fundamental problems underlie this shortfall. In general psychology, and to a greater extent in applied psychology, many experiments and investigations have been conducted in order to resolve specific questions and problems. However, a fundamental theory has not been advanced, nor has a theoretical basis been developed; there has been only empirical generalisation (Groeben and Westmeyer, 1975). At best, particular theories were produced which can only be applied to a very limited problem domain.
In view of the lack of a comprehensive model of traffic, or driver behaviour, it is possible to resort to theories from general psychology. This is, however, not very productive, as Huguenin has shown with an example of learning theories. Learning theories are simply insufficient for the explanation of specific behaviour in road traffic. Basically only the learning process is explained and the question answered of how certain types of driver behaviour could have originated, but they do not explain in which driver categories specific types of behaviour occur under certain conditions. As far as traffic psychology is concerned, the main interest lies with which part of the driver’s psychic inventory is reacting and how, and not with how that ability was acquired. For example, it can usually be assumed that a driver understands the use of the brake pedal because of his/her individual learning history. For traffic safety work it is, however, extremely important to know why that knowledge is incorrectly used in many cases. Theorists of learning may object that, in such cases, the learning process has not been finished; training with more variability of the learning situation and a better transfer possibility would eliminate such inadequacies. This answer seems, however, not only to bypass reality, but also to neglect aspects of cognition, motivation and theories of action (Huguenin, 1988).

Motivational models of driving

Motivational models of driving emerged in the 1960s and 1970s as alternatives to the skill-based models that had existed prior to that time (Summala, 1985). The main assumptions of these models are that driving is self-paced and that drivers select the amount of risk they are willing to tolerate in any given situation. The emphasis on transient or situation-specific factors come about in response to the lack of success of earlier attempts to relate stable personality traits to accident causation (Johnston and Perry, 1980). The risk associated with possible outcomes is seen as the main factor influencing behaviour, however, these models also assume that drivers do not generally make a conscious analysis of the risks associated with alternative outcomes (Wagenaar, 1992).

The importance of motivation in driving was demonstrated from a series of studies which have been conducted in Sweden. In the first studies (Johansson and Rumar, 1966; Johansson and Backlund, 1970), drivers were stopped immediately after they have passed a sign and asked to identify the last sign they have seen. Accuracy of reporting the sign under these conditions ranged between 17% and 78%, depending upon the “subjective importance” of the sign, that is, the amount of risk involved in ignoring the sign. In a subsequent study (Summala and Näätänen, 1974), an experimenter inside the vehicle asked drivers to recall all signs (581) over a distance of 257 km,
after driving in heavy traffic. Only 3% of the signs were ignored. Although methodological differences between the two studies preclude strong conclusions about their differences, the results have been cited repeatedly as evidence that on-road driving differs considerably from drivers’ capabilities, which in turn supports the conclusion that motivation is an important determinant of on-road driving (Näätänen and Summala, 1976).

Examples of motivational models include risk compensation models (Wilde, 1982), risk threshold models (Näätänen and Summala, 1976) and risk-avoidance models (Fuller, 1984). Risk compensation models propose a general compensatory mechanism whereby drivers adjust their driving (for example, speed) to establish a balance between what happens on the road and their level of acceptable subjective risk. Wilde’s risk homeostasis theory (RHT) is based on the assumption that the level of accepted subjective risk is a relatively stable personal parameter. An important implication of this model that drivers will compensate for traffic safety improvements by driving faster and/or less cautious to re-establish a constant level of risk. As a result, changes to the roadway or vehicle or even improvements to driving competency will not have a lasting safety impact. This implication has created considerable controversy (McKenna, 1988; Wilde, 1988; Evans, 1991).

Risk-threshold models propose the existence of a control process by which drivers attempt to maintain a stable balance between subjective, perceived risk and objective risk. The motivational model of Näätänen and Summala (1976), later renamed the zero-risk model (Summala, 1985, 1988) is of this type. According to their model, the perceived risk (R) in traffic is the product of the level of subjective probability of a hazardous event and the subjective importance of the consequences of the event. Behaviour is assumed to be directly related to the level of R. In most circumstances, R is perceived to equal zero, that is, drivers generally feel and act as if there is no real risk at all. If a threshold is exceeded, risk-compensation mechanisms are activated in an attempt to lower the risk level. The main differences between this and risk compensation models are the existence of a threshold and the operation of safety margins (Summala, 1988). Whereas, according to the risk compensation models, drivers are always adjusting their performance, the risk-threshold model assumes that compensation begins only when the perceived risk exceeds a threshold. Safety margins, defined in terms of the spatial or temporal distance between the driver’s vehicle and a hazard, are proposed as alternatives to the stable-risk parameter proposed by Wilde (Summala, 1988).
The risk-avoidance model (Fuller, 1988) is based on the assumption that making progress towards a destination and avoiding hazards are the two predominant driver motivations. The conflict between these two motivations forms the conceptual basis for this model. Focusing on avoidance of threats derives from the fact that it is impossible to drive unimpeded in a straight line to one’s destination, having to repeatedly avoid obstacles and potential hazards along the way. Repeated exposure to obstacles is the basis for learning how to identify risks on the road.

Motivational models have been criticised for lacking specificity regarding their internal mechanisms, which precludes validation (Michon, 1985; Molen and Botticher, 1988). In an attempt to overcome this problem, Molen and Botticher (1988) developed a hierarchical risk model, which they argue is both fully specified to allow quantitative calculations and flexible enough to be consistent with the three main risk models mentioned. However, their model has been criticised for failing to distinguish between the aggregate and individual levels of analysis, a problem that according to Michon (1989) will lead to “vicious circles and pernicious homunculi,” among other theoretical problems. Wilde’s RHT model suffers from the same problem, as evidenced by the need to assert that all drivers have homeostatic mechanisms to explain homeostatic behaviour at the aggregate level of analysis. Fuller’s threat-avoidance model does not suffer from this problem, however, as with all behaviourist models, it cannot handle embedded or nested behaviour, such as when a problem arises while the driver is dealing with another problem. Therefore, Fuller’s model is applicable only to single-instance situations (Michon, 1989).

The emphasis on risk has been criticised by those who argue that motivation is multifactoral. Rothengatter (1988) identified four different motives for speed selection, including pleasure in driving, traffic risks, driving time, and expenses. Summala (1988) identified a tendency towards higher speeds, reluctance to reduce speed, conservation of effort, and habit as motives in driving, citing the example of a driver who passes another just before exiting a road. Drivers may actually attempt to minimise their allocation of attention to driving, to free up resources for non-driving-related activities. The proliferation of in-vehicle distractions, including entertainment systems, communications systems, and even the increased incidence of reading while driving provide support for operation of this motive.

Janssen and Tenkink (1988) attempted to facilitate reconciliation between the RHT and its opponents by suggesting three modifications to the assumptions of that model. First, to address concerns over the existence of a very specific target level of accident risk, they argue that risk taking must be considered as part of a more general utility-maximising process. They also suggest
softening the strong requirements of RHT that compensation need be complete. Finally, they argue that risk taking must be modelled at different levels, including the trip and particular situations. Their model reflected the emergence of hierarchical control.

Motivational models focus on what the driver actually does in a given traffic situation rather than on the level of driving skill. The driver is seen as an active decision maker or information seeker (Gibson, 1966), rather than the passive responder implicit in many information-processing models. A related distinction has been made between performance and behaviour (Näätänen and Summala, 1976; Shinar, 1978; Evans, 1991), where performance refers to what drivers are capable of doing, while behaviour refers to what drivers actually do on the road. This distinction helps clarify differences between the major research paradigms used to study driving behaviour. Individual-differences research has focused almost entirely on predicting accident rates. To the extent that this research has used performance limits on information-processing tasks as predictors, it implicitly assumes that pre-crash behaviours represent the limiting capabilities of drivers. The questionably validity of this assumption and the restricted focus on the set of behaviours that precipitate crashes are likely reasons for the lack of success of efforts to identify predictors of safe driving. In contrast, motivational models address driving in its entirety and emphasise the inherent variability in driving. The importance of crashes is replaced by situational variables, such as safety margins, which by definition reflect a difference between on-road driving and performance limits.

Despite their appeal and promise, motivational models have failed to generate a significant body of research findings. Several possible reasons exist (Ranney, 1994). First, the confusion between individual and aggregate levels of analysis has plagued some of these models (Michon, 1989). The result is an inability to generate testable predictions. Second, the protracted debate concerning the validity of the theory of risk homeostasis (Janssen and Tenkink, 1988) has stalled progress. Finally, empiricists have failed to come to grips with the implication of these models that if driving is determined largely by motives, goals, and expectations, it may be irrelevant to study driving in the laboratory, driving simulator, or closed course, where the fundamental element of the goal of the trip is removed (Duncan, 1990). Other differences between on-road driving and laboratory studies also exist and relate to peripheral speed information and absence of the simulation of longitudinal forces (Törnros, 1998). One recent exception is a study of driver behaviour on curves conducted by Wong and Nicholson (1992), which combines motivational theorising with an observational methodology.
Information-processing models

Gibson and Crooks’ (1938) field theory of driving suggests that drivers move through the driving environment guided by a field of safe travel. This field is specified by visual information, and the driver chooses a path through the driving environment that avoids obstacles that would impede locomotion. This field combined with both the distribution of attention and response selection provide a framework for explaining driver behaviour.

According to Senders et al. (1967), any section of road can be modelled as having a certain information rate built into it – that is, there are so many bits per length of road. The faster a driver traverses a portion of the road, the more bits per unit of time must be processed. Were the driver to see a road only at fixed intervals, uncertainty would develop about where he/she is on the road. If the intervals between observations were very long, then the accumulated uncertainty and the amount of information to be absorbed on the next observation would be greater. If the observation time itself were very short, the driver would be unable to completely reduce the uncertainty by absorbing the required amount of information. Drivers tend to drive to a limit which is determined by that point where the driver’s information processing capacity, either real or imagined, is matched by the information generation rate of the road, either real or estimated. In addition, familiarity with roadway and traffic conditions reduces uncertainty and information flow rate, permitting a higher vehicle speed.

Information-processing models began to emerge in the 1950s (for example, Broadbent, 1958) in response to the discovery of air-traffic controllers’ limitations in handling simultaneous messages (Kahneman and Treisman, 1984). These models are typically represented as a sequence of stages, which include perception, decision and response selection, and response execution. Each stage is assumed to perform some transformation of data and to take some time for its completion (Wickens, 1992). Much experimentation has been directed at determining which types of processing can occur simultaneously and which must occur sequentially.

With several exceptions (Rockwell, 1972; Shinar, 1978), models of driving behaviour have failed to incorporate theoretical advances in cognition (Michon, 1985). In contrast to Michon’s suggestion that human factors curricula are to blame, Ranney (1994) proposed that early information-processing models and their associated experimental techniques were incompatible with the requirements of complex tasks such as driving. Specifically, in a deliberate attempt to isolate perceptual processes from memory, psychologists created experimental techniques using stimuli
that were abstract, discontinuous, and only marginally real (Neisser, 1976). As a result, the spatial and temporal continuities of real objects, essential for describing continuous tasks such as driving, were eliminated. During the 1970s, a paradigm shift (Kuhn, 1962) took place in the study of attention (Kahneman and Treisman, 1984). The shift involved a move away from determining the limits of processing and locus of the attentional bottleneck. More recently, based in large part on theoretical advances by Schneider and Shiffrin (1977) (also Shiffrin and Schneider, 1977), research has been directed at determining the characteristics and conditions under which automaticity develops. This work has influenced research in human factors (Fisk et al., 1987), and also the theory of driving behaviour.

**Automaticity**

The importance of automaticity to driving is not a new idea, having been identified by Gibson and Crooks (1938) as a worthy research endeavour. Automaticity is characterised as fast, effortless processing, which develops following extended consistent practice (Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). It is contrasted with controlled processing, which refers to slow, serial, and effortful processing. Although early experimentation used very simple stimulus material and examined elementary processes such as detection and memory search (Schneider and Shiffrin, 1977), recent work has extended the automatic/controlled distinction, although not the explanatory mechanisms, to more complex behaviours (Fisk and Schneider, 1983). Schneider and Shiffrin's (1977) theory suggests that virtually all behaviours include components of controlled and automatic processing and that the relationship between the various components is constantly changing according to the type and quantity of practice.

Because driving involves a seemingly endless variety of situations, a model of driving behaviour must allow for the development of automaticity in the absence of the highly consistent stimulus conditions that can be presented in laboratory settings. Several studies have explored the development of automaticity in situations which are not entirely consistent. Fisk and Schneider (1984) found that automatic processing was not limited to tasks that are consistent from input to output. They varied the consistency of attending and the consistency of motor responding factorially in a multiple-frame detection paradigm and found that consistent attending produced a substantial improvement in search performance regardless of the consistency of the response component. Automatic component processing can thus improve total task performance, despite inconsistencies among other task components. In the context of driving, this suggests, for example,
that braking and steering patterns may become automatised despite differences in the characteristics of precipitating situations, such as obstacles or hazards.

Fisk et al. (1988) examined the relationship between higher and lower level consistency. Using a magnitude-judgment task, they found that stimulus-based consistency is not necessary for automatic process development if higher order consistencies can be identified and used by subjects. Specifically, when the task was to identify the largest digit in a display, there would be no consistent relationship between any single digit and response. However, subjects were able to take advantage of the consistent higher-level relationships among digits, (for example, 7 is always greater than 6) for development of automaticity. Applied to driving, this may suggest that consistent practice following the same route to a destination can result in automaticity with regard to route selection, independent of day-to-day variations in weather, visibility or traffic conditions. Alternatively, the similarities among curves or intersections may be sufficient for development of automatic action patterns, despite geometric differences between individual intersections or curves.

Automaticity has also been used to refer to the detection of objects that can occur preattentively, without repeated exposure to consistent stimuli. Treisman and Gelade (1980) have proposed that features such as colour, orientation, size, and direction of movement are coded automatically, without attention. Furthermore, the relation between a target and its surroundings determines the ease of object detection. If an object differs from its background by the presence of a unique feature, it can be detected preattentively and the object appears to “pop out” of the background automatically. However, if the object differs by more than one feature, focused attention, which requires conscious effort, is necessary for detection. This model has not been applied to driving, in which the relationship between objects and their background is constantly changing, however Rumar (1990) has suggested that the artificiality of the driving environment and the unnatural speeds prevent drivers from taking full advantage of automatic detection capabilities that allow for efficient visual search in natural environments.

While motivational models of driving have been criticised for lacking detail concerning mechanisms, information-processing models have generally been criticised for not incorporating motivational or emotional components (Eysenck, 1982). Modelling the driving task would thus appear to offer an ideal forum for combining concepts from these two areas. For example, it is of interest to determine the relation between the various motives that influence driving and the development of automaticity. Summala (1988) discusses the relationship between uncertainty and the development of automaticity in driving, suggesting that novice drivers initially feel a sense of
uncertainty in most situations. With practice, skills become automatised and self-confidence replaces uncertainty. In driving, novel or hazardous situations evoke uncertainty, which, according to Summala (1988), causes control to shift from automatic to controlled, conscious processing.

Multiple resources

Because driving is a time-shared activity, a theory of driving behaviour should provide some basis for determining which combinations of skills can and do become automatised with practice. For example, basic vehicle control activities (steering, acceleration, shifting, braking) can be combined with visual search, decision making at intersections, reading traffic signs, listening to the radio, and even talking with a passenger or operating a telephone. Wickens’ (1984) multiple-resource theory may provide a framework for determining the degree of compatibility among various component tasks. He proposed the existence of several different supplies of resources, including the stage of processing (early, late), the modality (auditory, visual), and the processing code (spatial, verbal). Wickens has demonstrated that interference in a dual-task situation will be more likely when the individual tasks draw on the same pool of processing resources.

Hierarchical Control Models

Rasmussen (1987) differentiated among skill-based, rule-based and knowledge-based behaviours. Skill-based behaviour is the lowest level and involves automated schemata, consisting of well-learned procedures. Rule-based behaviour involves automated activation of rules or productions. Knowledge-based behaviour involves conscious problem solving and is generally invoked in novel situations for which no existing rules are applicable. Recently, Lehto (1991) proposed a fourth level, referred to as judgment-based behaviour. Schneider and Shiffrin’s (1977) distinction between automatic and controlled processing appears very similar to that between skill-based and knowledge-based processing (Rasmussen 1987), however, Rasmussen’s taxonomy apparently intends no dynamic relations between the different types of processing. Reason’s (1990) generic error modelling system (GEMS), described in detail in a subsequent section, has integrated processing mechanisms with Rasmussen’s model, thereby extending its use beyond classification of errors.

A three-level hierarchy has also been proposed to underlie cognitive control of driving (Michon, 1985; Molen and Botticher, 1987). The three levels include the strategic, tactical or manoeuvring, and operational or vehicle control. The strategic level involves general trip planning, including
setting trip goals (for example, minimise time, avoid traffic), selecting routes, and evaluating the costs and risks associated with alternative trips. The manoeuvring level involves negotiation of common driving situations such as curves and intersections, gap acceptance in overtaking or entering the traffic stream, and obstacle avoidance. The operational level consists of the immediate vehicle control inputs, which are largely automatic action patterns (for example, steering, braking, shifting). This hierarchy assumes a dynamic relationship among concurrent activities at the three levels; however, control mechanisms have not been specified.

The different levels of decision-making require different types of information. While strategic decision-making can be largely memory-driven, requiring little if any new information, manoeuvring and vehicle-control decisions are based on the immediate driving environment and can thus be considered as mainly data-driven (Norman and Bobrow, 1975). Another difference concerns the time available to make decisions. Decisions at the strategic level are generally not constrained by real time. General trip plans can be made in advance of a trip. More specific strategic decision-making can generally be done while driving as time permits, often many minutes before execution. Manoeuvre-level decisions are considered to take place in seconds, while control decisions require only milliseconds to execute.

The control hierarchy of driving has been related to Rasmussen’s taxonomy, as shown if Figure A.2 (Hale et al., 1990; Molen and Botticher, 1987).

<table>
<thead>
<tr>
<th>Strategic</th>
<th>Tactical/Manoeuvring</th>
<th>Operational/Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>Navigating in unfamiliar area</td>
<td>Controlling skid</td>
</tr>
<tr>
<td>Rule</td>
<td>Choice between familiar routes</td>
<td>Passing other vehicles</td>
</tr>
<tr>
<td>Skill</td>
<td>Route used for daily commute</td>
<td>Negotiating familiar intersection</td>
</tr>
</tbody>
</table>

**Figure A.2** Classification of selected driving tasks.

For experienced drivers, most driving tasks cluster in the three cells on the diagonal that runs from the upper left to the lower right box. Skill-based behaviour is involved at the operational level, rule-based behaviour at the tactical level, and knowledge-based behaviour at the strategic level. As shown by the examples in other matrix cells, exceptions reflect differences between skilled and novice performance, and between familiar and unfamiliar situations. For example, novice drivers
initially use knowledge-based behaviour to shift gears, while experienced drivers use skill-based, automatic action patterns. Experienced drivers can generally use skill-based behaviour for navigating along highly familiar routes or for negotiating familiar intersections, reflecting the fact that automaticity can operate at all levels of control. However, rule-based behaviour will predominate in unfamiliar situations, as long as previous experience is sufficient for selection of rules. Novel or unexpected situations, for which no applicable rules can be located, will disrupt skill-based (automatic) processing and necessitate knowledge-based (controlled) processing. In general, drivers operate more homogeneously and predictably at the skill-based and rule-based levels than at the knowledge-based level (Hale et al., 1990).

The incorporation of a hierarchical structure, together with the inclusion of mechanisms that enable control to switch between levels, which were proposed by Michon (1985) as criteria for a comprehensive model of driving behaviour, have provided new impetus for modelling efforts. Specifically, the conceptualisation of driving as concurrent activity at three different levels suggests that the driver's motivation may also have components relating to different levels of control. Motives concerning the purpose and importance of a trip may influence behaviour throughout the trip, however, situations encountered en route may create shorter-term goals that motivate tactical problem solving. For example, although a driver may have selected a route and departure time to ensure a leisurely, relatively uneventful drive, the presence of excessively slow traffic ahead may motivate the driver to speed up and pass. Compensatory behaviour may also operate at different levels of control. For example, changes in trip plans, such as the avoidance of rush-hour or nighttime driving by older drivers (Planek and Fowler, 1971), are examples of strategic-level compensations. Adjustments to safety margins, such as the rejection of a higher percentage of gaps during on-road merging by older drivers (Wolffelaar et al., 1987), or during conditions of poor visibility, are manoeuvre-level compensations. Momentary adjustments to steering and acceleration in response to slippery roads are examples of compensation at the vehicle-control level.

Motivational models assume that drivers actively decide how to allocate processing resources among the concurrent activities at the strategic, tactical, and operational levels of control (Michon, 1989). As shown above, behaviour at all levels may become automatic in highly familiar situations. However, changes in motivation, created by unanticipated deviations from the driver's expectations, are likely to disrupt automatic processing. Uncertainty, created by an unexpected event or associated with a conflict between motives at different levels of control, has been proposed as the mechanism that triggers compensatory behaviour, which leads to a reallocation of cognitive resources. However, the extent to which such decision-making occurs automatically, and thus
outside the driver’s awareness, is not known. Because the vast majority of behaviour occurs at the skill-based and rule-based levels, it is likely that drivers are rarely conscious of their decision making until knowledge-based problem solving is required (Wagenaar, 1992).

Errors

The contribution of errors to crash causation has been studied extensively. Perhaps the most widely referenced study is the Indiana University Study of Accident Causes (Treat et al., 1977). For this study, a taxonomy was developed allowing classification of causal factors as human, vehicular, or environmental. Human causes were further classified as either direct or indirect. Direct causes referred to errors that immediately precipitated a crash, while indirect causes referred to conditions such as fatigue, alcohol intoxication, or emotional upset. The results of this study have been cited extensively to support the importance of information-processing failure to crash causation, however, the direct causes precipitating a crash do not necessarily implicate a particular information-processing mechanism (Shinar, 1993). For example, a “decision error”, which might be associated with passing or crossing a traffic stream, could represent the outcome of any of a number of problems, including sensory deficiencies, misperception of critical information, application of an inappropriate rule, lack of appropriate knowledge, or perhaps a deliberate attempt to accommodate an extreme motive such as being in a hurry. The limitations of taxonomic models in providing information concerning underlying behavioural mechanisms have been discussed by Michon (1985).

Because drivers commit many errors other than those that precipitate accidents (Brown, 1990), it is clear that accident data alone do not provide appropriate information for the analysis of driving errors. In much the same way that motivational models have shifted the emphasis from accident-precipitating behaviours to all driving behaviour, more recent theories have considered errors as a part of normal behaviour (Ranney, 1994). This alternative approach advocates studying errors within the larger context of all driving behaviour, because they are inevitable in self-regulating systems (Rasmussen, 1990).

Drivers adopt safety margins to protect against the consequences of their errors. Effective safety margins require protection from the entire distribution of responses. The occurrence of accidents in relation to increasing speed and speed variability suggests that drivers’ safety margins are inadequate (Brehmer, 1990). Inadequate safety margins occur because drivers underestimate traffic hazards and overestimate their own driving skills (Brown, 1990). Brehmer’s model leads to
predictions concerning the types of errors to expect in various situations. For example, if drivers follow too closely, they will impose a time-stress on themselves and not allow sufficient time visually to sample the wide range of spatially distributed cues necessary for driving. A relatively high frequency of attentional and perceptual errors will likely occur as a result.

Brown (1990) distinguished between factors that influence the production of errors and those that constrain drivers' ability to recover from errors. Because of the potential for catastrophic consequences resulting from driver errors, the highway system has been designed to be tolerant of minor errors, such as deviations from the travel lane. The absence of feedback concerning minor errors in driving can weaken associations between actions and their consequences, which can lead to over-learning of inappropriate behaviours. Drivers' adaptation to error is thus prone to distortion, which may affect the degree to which correct responses can be automatized (Groeger, 1990). Instead, automatic action patterns may include a relatively wide range of both correct and incorrect responses. One implication is that drivers will be unaware of inappropriate speeds or inadequate safety margins. The inherent variability of human behaviour combined with the variability of automatic patterns will inevitably lead to more serious errors. At this point, the driver's ability to recover from error may determine the likelihood of accident. This has led Brown (1990) to conclude that factors that influence drivers' ability to recover from errors may be more important to theories of driving behaviour than factors that influence error production.

Visual search modelling

Visual search in driving involves identifying salient information in a constantly changing, moving scene (Ranney, 1994). Targets differ along dimensions of their familiarity, the predictability of their location, and their movement. For example, regulatory signs and signals occur in fairly predictable locations and contain information that is highly predictable. In contrast, dynamic information displays, such as changeable message signs, may be located at less predictable locations (for example, construction zones) and may contain less predictable information (for example, temporary speed limit, lane closed). Moving targets include other vehicles, pedestrians, and unexpected hazards, such as debris falling from a vehicle or a rolling ball (likely to be followed by a child).

Visual search paradigms in the laboratory have been used to study mechanisms of selective attention (Shiffrin, 1988). According to Theeuwes (1989), the essential difference between visual search, as studied in the laboratory and in driving, concerns the existence and nature of the search
targets. In driving, the search targets may not be well defined and search is not top-down (Norman and Bobrow, 1975) or target-driven, but rather bottom-up or data-driven. In other words, much of visual search in driving consists of drivers’ waiting to notice a conspicuous target (Cole and Hughes, 1990). In one sense, this may render much of the visual search literature not applicable, because most experimentation on selective attention has used top-down, target-specified search (Johnston and Dark, 1986). Other differences between laboratory studies and on-road driving include the use of eye and head movements while driving and the constantly changing visual scene in the moving vehicle. The majority of laboratory studies use static displays with minimal content, to minimise the involvement of memory, and brief exposures to eliminate eye movements.

These differences are considered in a model of visual search in driving developed by Theeuwes (1989). According to this model, top-down regulated search during driving occurs only when several conditions are met. First, the driver can be in two possible states, a state of certainty or a state of uncertainty. This state is determined by the changing sensory input in relation to the immediate goals of the driver. For example, if the driver expects to stay on the same road for sometime, he/she will be in a state of certainty with regard to the question of, for example, where to turn. If the driver is near a point where a turn is required, in an unfamiliar area, then he/she may become uncertain. If there is no uncertainty, then no search target will be generated and the driver will be passively noticing, rather than searching. However, when the driver becomes uncertain, an attempt will be made to reduce the uncertainty. The type of uncertainty, together with the outcome of a global analysis of the environment and the driver’s experience with this type of uncertainty in this environment, will determine whether a search is initiated. If a search objective is defined, a subsequent process will determine whether an appropriate schema exists so that a learned stereotypic search pattern can be used. For example, the uncertain driver looking for a place to turn may activate a schema that directs his/her gaze to the location most likely to contain a street-name sign or a break in the pavement edge-lines. The two-stage search model, in which an initial global search is followed either by a selected specific scanning pattern or a more general-purpose pattern, is based on Rabbitt’s work (1981, 1982).

This model applies the mechanisms of motivational models to visual search in driving. Specifically, the control of search by an uncertainty-reducing mechanism is consistent with the risk-threshold model (Naätänen and Summala, 1976). Furthermore, the emphasis on conspicuity provides the framework for application of models of visual detection. However, Ranney (1994) is of the opinion that additional theoretical work will be necessary to determine the extent to which the automatic detection of conspicuous targets in driving occurs as the result of extensive learning of
consistently mapped relations (Schneider and Shiffrin, 1977), or through the action of distinctive features (Treisman and Gelade, 1980).

Models of cognitive processing and errors

Reason's (1990) generic error-modelling system (GEMS) has incorporated information-processing mechanisms into Rasmussen's taxonomy in an attempt to show how control shifts between levels. The model concerns two types of error, including monitoring failures, which precede the detection of a problem and problem-solving failures, which follow such detection. Periodic attentional checks are an important part of well-practised (skill-based) actions. These checks are intended to determine whether the actions are running according to plan and whether the plan is still adequate to achieve the desired outcome. The scheduling of attentional checks can be a critical factor contributing to the occurrence of a monitoring failure. Attentional checks should occur near critical choice points, particularly if the planned action is not the most frequently used choice. For example, if a driver selects a route that corresponds initially to a highly familiar and frequently used route, but later requires a change, the likelihood of an error may depend on whether an attentional check occurs slightly before the point at which the driver must deviate from the highly practised route. Reason (1990) has identified two categories of monitoring failure, including inattention (distraction) and over-attention (preoccupation).

Problem-solving failures result from attempts to apply inappropriate rules. This derives from the assumption that human problem-solvers are strongly biased towards looking for an existing solution at the rule-based level before resorting to the considerably more effortful knowledge-based solution. In fact, Reason suggests that even when an appropriate rule cannot be found, the bias towards finding an existing solution motivates continued attempts to identify similarities between the current and previous situation, even as knowledge-based problem-solving proceeds.

Consistent with Michon's (1985) criteria for a model of driving behaviour, GEMS includes the provision for switching control among the various levels. Once an attentional check results in the detection of a problem, control will shift from the skill-based to rule-based level. This change can also be represented as a change from primarily automatic to controlled processing. If an applicable rule can be found, it will be activated, however, control will remain at the rule-based level until it has been decided that the new rule will resolve the problem. At that point, control will again shift to the skill-based level. However, if no applicable rule can be found to address the immediate problem, control will eventually shift from the rule-based to the knowledge-based level. According
to Reason, this occurs when the problem solver becomes aware that none of the existing rules is applicable to the current problem.

This model represents a combination of cognitive processing mechanisms and hierarchical control theory. Specifically, automatic versus controlled processing in interpreted within the context of Rasmussen's three-level control hierarchy. This model is also consistent with motivational models in that subjective uncertainty is viewed as the mechanism that triggers a shift in the allocation of attentional resources.

Ranney noted that attentional mechanisms have been prominent in all approaches to understanding driving behaviour. However, whether crashes are more often caused by systematic errors associated with drivers' performance limits, such as one's ability to switch attention rapidly among competing sources, or by variable errors of automatic processes, such as lapses of attention, is not well established. The complexity of attentional mechanisms and lack of operational definition underscore the need for a common terminology for attentional and control mechanisms in driving. Furthermore, Ranney noted that the framework provided by the hierarchical control model, together with concepts derived from work in automaticity, appear to be useful in this endeavour.

A.7.3 Transport automation and driver behaviour modelling

There are a number of reasons why in recent years electronic driving aids were developed and implemented at an increasing rate (Brookhuis and Van der Heijden, 1999). The first reason is comfort (a good selling argument), but also economic principles (like efficiency) are a compelling drive, while environmental arguments play a role of increasing importance. Finally, yet importantly, safety (i.e. the unacceptable number of accidents) is an important issue in this respect.

Driver comfort appears to be a strong impetus for the development of electronic driving aids, at least from a marketing point of view. Car manufacturers are keen on driver comfort and invest considerable effort in the development and improvement of comfort enhancing electronic aids. Well-known examples of this type of applications are navigation or route guidance systems and advanced cruise control systems. Though expensive, prototypes of this type of systems passed a number of tests on functionality, efficiency and validity, where after improved versions were successfully placed on the consumer market. Before the actual marketing, user needs research (or marketing research) is indispensable, but also studies on acceptance and certainly behavioural effects are still necessary after implementation to ensure and consolidate market penetration and
safety. Stakeholder acceptance is dependent upon such requirements as system safety, reliability, functionality and validity (does the system function correctly), and benefit (is there a positive cost-benefit balance). Finally, environmental issues are not decisive in this area yet, but will gain weight in the future.

The prevention or reduction of traffic accidents requires countermeasures that have to be devised and introduced to prevent those behaviours contributing to accidents. In Europe, the USA and in Japan, combined ergonomic and engineering approaches to both hazard assessment and the indication of drivers’ performance limits have developed into research and development of new and relevant primary safety measures to prevent accidents. Secondary safety measures are specifically designed to prevent or reduce injuries once a collision has become inevitable. Brookhuis and Brown (1992) argue that an ergonomic approach to behavioural change via engineering measures leading to transport automation in the form of electronic driving aids, needs to be adopted in order to improve primary safety, transport efficiency and environmental quality.

Transport automation

Transport automation in general pertains to all kinds of electronic driving aids for all kinds of applications in traffic and transport. Electronic driving aids may operate in advisory, semi-automatic or automatic mode (Rosengren, 1995). The latter type has only recently been introduced in society. An example of semi-automatic driving aids is the ISA (Intelligent Speed Adapter, see Brookhuis and De Waard, 1999). An ISA takes into account the local speed restriction and adjusts the (maximum) driving speed of the vehicle to the posted speed limit.

Transport automation, or what is now called ADAS (Advanced Driver Assistance Systems), is to be considered as the collection of systems and subsystems on the way to a fully automated highway system. Only when, on a fully automated traffic lane, the vehicle can be operated under fully automated control, which is very similar to the automatic pilot in aeroplanes (Congress, 1994), the human factor can be bailed out. ADAS concepts include among others blind spot detectors, Adaptive Cruise Control (ACC), Autonomous Intelligent Cruise Control (AICC), and platoon driving.

ADAS has a considerable history. In Europe, several car manufacturers and research institutes started the PROMETHEUS initiative, around 1986. A series of projects was carried out under this umbrella, most of them aiming at practical solutions to urban traffic problems. The European
Union initiated the DRIVE (Dedicated Road Infrastructure for Vehicle safety in Europe) programme shortly thereafter, in which a considerable number of projects tackled practical problems as well as fundamental issues. An example of the latter is the GIDS (Generic Intelligent Driver Support) project, the largest project in DRIVE, ahead of its time and still relevant (Michon, 1993). The overall objective of this ambitious project was to determine the requirements and design standards for a class of intelligent driver support systems which will conform to the information requirements and performance capabilities of the individual drivers. On the one hand this class of systems will aid the driver’s detection and assessment of road and traffic hazards, on the other they will provide guidance onto the driver’s ability to deal with specific hazards.

Problems with ADAS

However, there are also potential problems to be expected with ADAS. For instance, increased complexity of the “cockpit” increases the likelihood of failure by the driver, and of at least one of the system’s components, either by spontaneous failure or by design error (Janssen et al., 1992). System failure requires additional, in fact continuous alertness of the driver, while at the same time driver alertness in general and attention for the driving task per se is decreasing in case of automation of the driving task (Brookhuis and De Waard, 1993). De Waard et al. (1999) found that in case of active take-over of vehicle control, only 50% of all drivers responded when the system failed. A few studies found support for over-reliance on automated (ADAS) systems, while others reported deterioration in driving performance (Brookhuis and Van der Heijden, 1999). This and other forms of behavioural adaptation, or compensation as it is called in a wider field, are factors that should be taken into account as well, when investigating the conditions for introduction of ADAS (Verwey et al., 1996).

Complacency

When a system fails to work or is in a state where failure is possible, feedback should be provided in order to let the driver know that he/she cannot rely on the system. The main reason for this is that automated systems can and will lead to complacency (Wiener and Curry, 1980). Complacency is an attitude of (over-) reliance on an automated system. In a test of reaction time to a system failure cue, Knapp and Vardaman (1991) found support for complacency, i.e. the reaction time to the cue increased compared to normal task performance. Ward et al. (1995) also found evidence for complacency, i.e. poor lane position control and failure to yield to other traffic was more frequently observed in drivers driving a car with AICC compared to drivers driving a normal car.
Therefore, in the full automatic highway system (AHS), complacency should be expected, as found in a simulator experiment, reported by De Waard et al. (1999).

**Human supervision**

The classic goal of automation is to replace human manual control, planning and problem solving by automatic devices. However, these systems still need human beings for supervision and adjustment. It has been suggested that the more advanced a control system the more crucial is the contribution of the human operator (Bainbridge, 1983). The point made by Bainbridge (1983) is as follows: normal operation is performed automatically; abnormal conditions are to be dealt with manually. Unfortunately, as a result of automation, experience is limited, while in case of abnormal conditions (i.e. something is wrong with the process) unusual actions will be required. Also, human problem solving is not optimal under time-pressure. Monitoring of (present) automatic processes is based on skills that formerly manual operators had, and that future generations of operators (drivers) cannot be expected to have (Bainbridge, 1983). Pilots also indicated that although automation reduced workload, it also had a negative effect on flying skills. They considered manually flying a part of every trip important to maintain these skills (McClumpha et al., 1991).

**Behavioural adaptation**

Automation may increase reaction time (Brookhuis and Van der Heijden, 1999). In case of continuous monitoring, reaction time to events in a driving task can be restricted to about one second, while if more functions have to be monitored and other task are attended to, awareness of the situation has to be refreshed with increase frequency. At the same time, possible malfunction and its origin have to be determined which might take many seconds. Automation is an attempt to reduce workload, but it is actually very likely to lead to increased workload (Hancock and Parasuraman, 1992) and changes in strategies and tactics of the drivers, in other words behavioural change or adaptation. There are several studies that have shown that monitoring of systems for malfunctions during prolonged periods of time induce high levels of workload, despite the fact that information processing requirements for these tasks are low in itself. Humans are poor ‘process monitors’ (for example, Molloy and Parasuraman, 1996) and enforced vigilance in the operational environment is very stressful (Hancock and Parasuraman, 1992).
Safety effects

The following mechanisms whereby ADAS applications can affect traffic safety were distinguished by Draskóczy (1993):

a) Direct safety effects due to changes in workload and distraction,

b) Increased reliance on information systems resulting in a reduction of driver's own observations, i.e. paying less attention,

c) Behavioural adaptation effects in the sense of changing driving behaviour, for instance, changing speed,

d) Modifications of accident consequences,

e) Altered exposure due to frequency and length of travel, modal choice, and route choice.

These mechanisms allow safety effects in either direction. While some systems are likely to contribute positively to traffic safety, there is reason for suspicion that other systems might result in reduced safety (Verwey et al., 1996).

Functionality and validity

When an electronic application is bought and installed in a motor vehicle, the operational characteristics are expected by the user to be according to their cognitive representation of the device. The cognitive representation of the functionality of electronic applications is highly dependent upon experience of the (potential) user, be it from comparable devices, similar conditions, hearsay, demonstrations, or literature in one form or the other. The validity of the functionality is therefore as much a subjective matter as the purpose the developer or manufacturer originally had in mind at the moment of marketing the device.

Modelling driving behaviour

In order to put maximum effort in avoiding the expected problems with transport automation or ADAS (Advanced Driver Assistance Systems), adaptive interfacing with the aim to keep the driver just sufficiently in the loop, should be investigated and put to test in a simulator and in real traffic, in an instrumented vehicle (Brookhuis and Van der Heijden, 1999).

The mentioned research questions that will be studied in the BATA programme (Brookhuis and Van der Heijden, 1999), can only be dealt with in an adequately flexible, controllable and fully
programmable environment. A driving simulator fulfils such a requirement in principle; however, the model that constitutes the traffic environment is still too simple. Simulated participants in this environment presently act according to a relative simple model of human behaviour in which perception, decision and action is perfect. Determinants of specific human characteristics in relation to the use of electronic applications can only be studied sufficiently realistic if the traffic environment responds sufficiently humanlike. A model in which human characteristics are to be expected includes human imperfections. Such a model for simulated ‘agents’ in a simulated traffic environment is still evolving. The BATA programme enables systematic research on HMI (Human Machine Interactions) and predictions about behaviour of the traffic environment to the operation of electronic applications in any degree of penetration. Additionally, the impact of transport automation on the behaviour of specific groups of traffic participants, such as the elderly, error-prone or inexperienced drivers, can be investigated (Brookhuis and Van der Heijden, 1999).

A.7.4 Example of an information-processing driver model based on artificial neural networks

Monitoring and warning systems in road vehicles have been proposed for a long time. However, they usually lacked driver acceptance because of inadequate (neither situation-specific nor driver-adapted) warning thresholds (Onken and Feraric, 1997). Michon (1993) gave a representative review of the investigations and development efforts in the past. However, none of them, including the GIDS (Generic Intelligent Driver Support) project described by Michon (1993), has gone beyond conceptual views for driver-adapted warning on the basis of the behaviour model of the actual person driving. Developments in new techniques for knowledge-processing have produced autonomous functional capabilities, like autonomous situation assessment, which can be exploited to support the driver’s situation awareness through situation-specific alert messages and warnings which are, at the same time, adapted to the individual demands of the driver. The Driver Assisting System (DAISY) has been developed to comply with these requirements (Onken and Feraric, 1997). The development was mainly conducted within the European PROMETHEUS project. In its first version of implementation, DAISY was designed for the freeway scenario. Further development would include the highway and urban roadway environments.

Rasmussen (1983) has categorised human cognitive behaviour into three main levels: knowledge-based, rule-based and skill-based behaviour. DAISY mainly has to deal with rule-based and skill-based behaviour of the driver. Therefore, it was one of the main design tasks for DAISY to develop a model of the driver for these behaviour levels and to achieve the ability to learn the individual driving behaviour on the skill-based level on-line.
To build a driver model an association between the actual traffic situation and the anticipated driver reaction has to be made. A model based on artificial neural nets can incorporate important characteristics of the driving task, for example, (i) driving behaviour is extremely situation specific, (ii) the driver continuously perceives lots of information which influences his/her behaviour, and (iii) driving behaviour is non-linear. To select a neural net paradigm for driver modelling in DAISY, technical requirements also had to be taken into account, for example, the learning module must have real-time processing capability, and gathered knowledge about driving behaviour must be stored (stability) while new knowledge must be incorporated in the system (plasticity). To meet all of these requirements Fuzzy ART, including necessary adaptations, was chosen for the modelling task (Onken and Feraric, 1997).

Daisy architecture

To account for the required situation adaptivity, DAISY contains a Situation Analysis Module (SA). This module performs an explicit analysis of the current objective traffic situation by classifying it from a driver’s point of view on the basis of environmental data, i.e. looking at situation features which are relevant with regard to the actual driver tasks.

The Model of the Actual Driver (MOAD) accounts for the adaptivity to the individual driver. In this module the actual driver’s normal driving style is extracted from observations of his/her behaviour during a learning phase. In MOAD the driver objectives (for example, desired driving speed) are determined first. Then the driver intent is recognised by use of a rule-based model. In the trained condition skill-based driver control inputs can be predicted, while the individual normal driving behaviour, described by characteristic time reserve values, can be extracted. A further processing step is the assessment of the current driver resources.

The Warning System (WS) combines the output of SA and MOAD to issue a situation-specific danger warning/alert in the case of dangerous deviations from normal behaviour caused, for example, by distraction, degraded vigilance or driver overload.

Adaptation of Fuzzy ART to the modelling task

The first adaptation step is to associate time series of driver reactions with classes of traffic situations or scenarios. Traffic scenarios are represented as feature vectors which include sensor
measurements of the vehicle itself, roadway parameters and information about other vehicles. Driver reaction is sampled and the signals are combined into vectors of time series of entire reaction sequences. To model the driving behaviour for lateral control, steering wheel rate, and for longitudinal control, brake and gas pedal positions are monitored. After classification of the situation-describing vector by the self-organising mechanisms of Fuzzy ART, a vector, containing sampled driver reactions, is associated with the detected class during the learning phase. This adaptation provides for the learning of driving behaviour by observation; this principle was used in a similar way but with a different neural network paradigm by Kraiss and Küttelwesch (1992).

The second adaptation step involves building a hierarchy of Fuzzy ART networks for classification of traffic scenarios to make provision for better model accuracy and to satisfy the requirement for real-time processing. Fuzzy ART networks on the first level perform a coarse clustering process. For every class found on the first level a second clustering process with higher resolution is made. Every traffic scenario class found for the second stage is then associated with the driver reaction sequences.

The third adaptation step is driven by the idea that different elements of the feature vector will have varying relevance in different traffic scenarios. Therefore a weighting of each element must be done. This is accomplished by introducing relevance parameters in the Fuzzy ART algorithm (Onken and Feraric, 1997).

Designing an artificial neural network to solve a specific problem also calls for the selection of network parameters by trial and error. The two vigilance parameters for coarse and fine clustering together with the introduced parameters for situation specific weighting of the input vector components must be tuned to get accurate driver models in DAISY. Genetic algorithms have proven to be a very flexible and a generally useable method to solve search and optimisation problems (Goldberg, 1989). This method was utilised to optimise the Fuzzy ART parameters off-line.

The trained driver model can be used in a driver assistance system like DAISY in different ways. To satisfy the requirement of situation-specific or driver-adapted messages and warnings it is possible to extract the driver and situation-specific time reserve values to describe normal driving behaviour on the basis of time. Deviations from normal driving can be detected and used to issue driver adaptive warnings. It is also possible to use a trained neural network as a humanlike
controller. Other possible applications like driver intent recognition, assessment of current driver resources and driver identification are part of the current research work (Onken and Feraric, 1997).

A.7.5 Example of an information-processing driver model based on pre-defined rules and algorithms

The driver model in the MIXIC software presented in Hogema (1999) was developed partly based on literature and partly on the author’s own experimental research with a driving simulator and an instrumented vehicle. The driver model includes free driving (speed control), car following (distance control), and lane-change behaviour. It interacts with a vehicle model at a detailed level (pedal and gear status). The driver model uses the vehicle controls with the aim to reach or maintain the individual, situation-dependent desired speed or following distance. The model includes aspects such as reaction time, perception thresholds for speed differences, and comfort levels for accelerating and decelerating. Since the traffic simulation model was designed to study the effects of Adaptive Cruise Control (ACC), the driver model also includes a component that describes the driver’s interaction with ACC.

When the vehicle has no ACC or when the ACC is switched off, the driver model sets the input to the vehicle model. For a vehicle fitted with ACC, the driver model determines when the ACC is activated and deactivated, taking into account the limited deceleration capabilities of ACC. After a driver has switched the ACC on, the vehicle inputs are determined by the ACC. The task of the driver then changes from actively controlling the vehicle’s longitudinal motion to monitoring the control behaviour of the ACC. When necessary, the driver model can disengage the ACC and thus take control.

Driver submodels

The driver model consists of three main components. The first two components describe the actual driving behaviour: a lane-change model and a longitudinal driving model. In each simulation time step, these are executed to produce the lane-change decision and the driver’s actions on the vehicle controls, respectively. The third component describes how the driver deals with the ACC. Driver model algorithms for different types of drivers have the same structure: they differ only in parameter settings. Differences, which exist among drivers, are implemented by means of sampling from (normal) distributions. In addition, there are a number of parameters that are assumed constant for all drivers.
Longitudinal driving model

The interface between the driver model and vehicle model has been defined at a detailed level, for example, positions for the accelerator/gas pedal, clutch pedal, and gear shift, together with the force exerted on the brake pedal are taken into account. The first step in the longitudinal driving model is to calculate the driver’s desired acceleration as a result of the current status of the vehicle and its relative position and speed to leading vehicles. Next, the desired acceleration is used together with the current status of the vehicle to determine the application of vehicle controls in such a way to achieve the desired acceleration as much as possible. This part of the model represents the “knowledge” a driver has about the vehicle’s reaction to pedal manipulations.

The longitudinal driving model distinguishes free driving and car-following behaviour. The desired acceleration is calculated continuously for both situations, and the most restrictive value is used as the resulting desired acceleration.

Free driving

If the current speed deviates more than a given proportion (set at 3%) from the desired speed, the desired acceleration is made proportional to the speed error, taking into account the reaction time of the driver.

Car following

In the car-following situation, the driver has to adjust his/her speed or following distance (or both) with respect to the traffic ahead. The core of the model implemented in MIXIC is derived from the optimal control model of Burnham et al. (1974). It is based on the assumption that the driver tries to keep the relative speed to the lead car zero and simultaneously attempts to keep the distance headway at a desired value.

The evaluation of headway and relative speeds takes the reaction time of the driver into account. Measurements of driver reaction times have been reported in the literature for a variety of situations, with values ranging from a few hundred milliseconds to several seconds (Triggs and Harris, 1982). In the context of MIXIC, driver reaction time represents the delay between a perceived deviation from the desired speed or headway and the beginning of the resulting correcting
foot movement. Delays for moving the foot between the accelerator and the brake pedal and perception thresholds are modelled separately, therefore, the MIXIC reaction time selected are smaller than most overall reaction times reported in the literature. Furthermore, since each driver’s reaction time is assumed constant over time, its value should be representative for dense traffic conditions, that is, with the driver alerted and prepared to react. For such conditions, Johansson and Rumar (1971) reported a median brake reaction time of 0.54s. Since this reaction time still includes a foot movement from the accelerator pedal to the brake pedal, a mean value of 0.3 sec and a standard deviation of 0.05 sec are used for the plain reaction time distribution in MIXIC. The time it takes a driver to move his/her foot from the accelerator pedal to the brake pedal has been reported in several studies (Snyder, 1976; Davies and Watts, 1969, 1970). It appears that the mean movement time lies in the range of 150-200 msec. Since MIXIC works in multiples of 0.1 sec, a value of 0.2 sec has been implemented for all driver types.

Lane-change model

The lane-change decisions are made based on a set of rules which evaluate the driver’s wish to carry out a lane-change manoeuvre and the possibility to carry out the manoeuvre. These rules contain the status of the longitudinal driving model: for example, they include a comparison of the current speed with the desired speed and with the speed of the lead vehicle in the current lane. Also, the vehicle’s actual acceleration with respect to the maximum comfortable acceleration and deceleration is evaluated. The rules for the manoeuvre possibility refer to the safety of the intended manoeuvre, determined by a sufficiently large gap in the target lane.

A.7.6 Autonomous driver model in driving simulation environments

Despite numerous issues about the validity of test results from driving simulations (Ranney, 1994), simulations are still regarded as an invaluable tool for conducting driving-related research (Papelis and Ahmad, 2001). In order to enhance the driver’s immersive experience in any virtual driving environment, autonomous traffic is necessary. The higher the fidelity of a driving simulator, the more stringent are the requirements for the realistic appearance of autonomous traffic. Unfortunately, an increase in the fidelity, or the complexity, of the traffic simulation model does not always enhance the appearance of realism. Papelis and Ahmad (2001) give an overview of a comprehensive autonomous driver model which is part of a microscopic traffic simulation system used in simulator applications at the National Advanced Driving Simulator (NADS).
In their experience, users generally focus their evaluation of the model realism on the richness of the behaviours, not the fidelity of the behaviours. When observing an autonomous driver model, a person is more likely to observe that vehicles did not overtake when they should have or did not move over to avoid a collision with an oncoming vehicle, than that the following distance did not vary according to a validated model. To some degree, this can be explained by considering observability of these behaviours. Many catastrophic problems in traffic simulation systems, such as deadlocks, collisions, or infinite wait situations, can be attributed to the lack of behaviours rather than to low fidelity or models that have not been validated.

However, Papelis and Ahmad warn that they are not implying that developing high-fidelity models of behaviours or validating traffic models is of no use in driving simulator applications. To the contrary, certain types of research necessitate validating at least part of the model. However, it is important that, in general, the range of behaviours is as, and sometimes more, important than their fidelity. After a set of minimal behaviours, has been developed, validation of individual behaviours is the next step in the evolution of the model.

The autonomous model described in their paper accommodates a large number of individual behaviours (called children), including the following: lane tracking, close-coupled object following, decoupled object following, lane change, oncoming traffic collision avoidance, static object collision avoidance, highway entry and exit ramp merging, side obstacle avoidance, slow or stopped vehicle overtaking, and an extensive set of intersection navigation behaviours.

The key component of the model is how the Autonomous Model fuses the outputs of all its children in determining the actual control inputs conveyed to the simulator. The fusion algorithm explicitly depends on certain mutual exclusion criteria among children and obeys the following priority rules:

a) The more conservative control inputs have higher priorities. Here, conservative indicates a lower desired speed or a target point that minimises the steering angle.
b) The Obstacle Avoidance has higher priority than Overtake.
c) Once the output of a specific behaviour (or child) has been selected, its output will maintain the higher priority until the manoeuvre has been completed.

Each individual behaviour (or child) is contained and described by an algorithm. For example, the Follow Behaviour (FH) provides a desired speed to ensure following a lead vehicle without collisions, and outputs desired speed, desired distance to lead vehicle, actual distance to lead vehicle, and the id of the lead vehicle.
The FH first identifies the lead vehicle among all vehicles located near its object. After obtaining the lead vehicle’s velocity and acceleration, the FH utilises a hybrid controller whose primary goal is to match the speed of the two vehicles, and whose secondary goal is to maintain a specific distance behind the lead vehicle. The desired speed for the two goals is calculated separately, and the final control input is a variable-weighted average of the two goals. The weights vary based on the time to collision with the lead vehicle. When the differential speed is high, the speed-matching controller is used almost exclusively; when the speeds of the lead and current vehicles match, the position controller is used almost exclusively.
A.8 Drivers' Emotions

A.8.1 Summary

Stress results when the perceived demands of the driving task are appraised as exceeding the driver's ability to cope with them. There are marked individual differences in self-reported susceptibility to stress during driving. Dimensions associated with different aspects of driver stress are aggression during driving, dislike of driving, alertness while driving and emotional reactions to overtaking and being overtaken. Individual differences in self-reported driver stress are related to individual differences in a variety of measures of personality, cognitive appraisal and emotion. General driver stress is associated with higher levels of neuroticism, aggression-hostility, frequency of everyday hassles, self-rated poor concentration and absentmindedness, and with stressed mood. Furthermore, involvement in automobile accidents is correlated with an insufficiency in the ability to control changes in emotion, emotional lability and excessive dependence on environmental stimuli.

Motivational factors determine what drivers do or what they must do with their skills. Among all motivational factors (transient motivation, personality and attitudes towards driving), it is crucial to know which skills are more sensitive to the effects of transient motivational factors. Drivers sometimes find themselves at the wheel following a happy or unhappy event related to their work or their social and family life. These factors can give rise to differences in the allocation of attention, changes in the choice of a decision (for example, choosing a more risky alternative), or in the carrying out of a decision (for example, a slower decision). These effects might be understood in the light of motivational consequences of loss of control. The stability of the effect of transient variations in control motivation in more experienced drivers appears to be an important aspect of driving experience. An important question is to find out whether the greater independence of the more experienced drivers to the motivational context (as compared with the novices) can be explained solely in terms of automatisation of information processing or whether one needs to consider specific regulation processes of motivational states.

One of the most relevant issues in applied social psychology research deals with the effectiveness of fear-arousing messages for changing attitudes and behaviours. A persuasive message presenting a person with the possibility that he/she is at risk with regard to the occurrence of an aversive event is effective in changing his/her behaviour. The fear aroused by a threatening message leads the person to mentally and behaviourally rehearse recommended actions to avoid the danger and thus
reduce the fear. Threat appeals have been used for many years as the main mean to foster careful driving-related attitudes and behaviours. These appeals usually consist of road-trauma films, showing crashed cars, bereaved parents, or injured people who were involved in car accidents. The goal of these appeals is to elicit fear and to remind people of their own vulnerability and finitude while driving in a reckless manner. However, there are opposing views on the effectiveness of driving-related threat appeals.

Driving has become intolerably stressful, dangerous, and demeaning. Drivers are stressed, threaten each other, are in a bad mood, terrorise their passengers, and often fantasise violent acts against each other. The development of a driving theory based on self-witnessing reports make it possible to construct a classification scheme or taxonomy to identify the components of driver behaviour from the perspective of the driver’s world. Such an inventory is useful for driver assessment and driver education and can provide norms or expectations of driving skills and errors in the affective (emotional), cognitive, and sensorimotor areas of behaviour. One needs to map the behaviour of drivers under varying social and psychological conditions to arrive at a comprehensive theory of driving behaviour. Feelings, thoughts, and perceptions are as much traffic and transportation issues as road conditions and traffic flow.

A.8.2 Personality, driver stress and accident proneness

Personality correlates of driver stress

Subjective driver stress may be explained by a transactional model (Lazarus and Folkman, 1984): stress results when the perceived demands of the driving task are appraised as exceeding the driver’s ability to cope with them. There are marked individual differences in self-reported susceptibility to stress during driving. Within the transactional model there are several causal routes through which personality and driver stress may come to be correlated. First, personality traits may be one of several causal factors which interact to determine stress reactions to traffic incidents. Since personality traits appear to influence appraisal (Martin, 1985) and cognitive coping strategies (McCrae and Costa, 1986), personality may directly affect the cognitive evaluation of traffic situations. Personality may also be associated with more low-level emotional and behavioural reactions to driving, associated with individual differences in neural function. For example, personality may be associated with individual differences in arousal, and hence preferred level of stimulation (Eysenck, 1967), or in sensitivity to signals of threat and reward (Gray, 1981). Second, personality may be associated with more generalised stress syndromes, perhaps provoked
by life events, which will find expression during both driving and other activities. For example, life events may evoke particularly high levels of stress in people high in neuroticism, which may in turn increase the stress experienced during driving. Third, personality may be directly affected by the person’s experiences as a driver. For example, prolonged exposure to driving on congested motorways could conceivably have adverse effects on personality. However, self-reported driver stress can only be expected to indicate general trends, which may not apply to some driving situations (Matthews et al., 1991).

Gulian et al. (1989) conducted two factor-analytic studies of self-reported driver stress. Five positively intercorrelated factorial dimensions associated with different aspects of driver stress were replicated across both factor analyses. These dimensions were associated with aggression during driving, dislike of driving, alertness while driving and emotional reactions to overtaking and being overtaken. Matthews et al. (1991) demonstrated that individual differences in self-reported driver stress are related to individual differences in a variety of measures of personality, cognitive appraisal and emotion. General driver stress is associated with higher levels of neuroticism, aggression-hostility, frequency of everyday hassles, self-rated poor concentration and absentmindedness, and, weakly, with stressed mood.

The most sharply discriminating scales of driver stress are Aggression and Dislike of Driving. Aggression appears to be associated most strongly with behavioural or social aspects of driver stress, in that this scale correlates with self-reports of accident history and of potentially dangerous driving behaviours (Gulian et al., 1989), but is only weakly related to self-appraised attentional efficiency and mood. These aggressive behaviours appear to be related to general personality characteristics associated with negative interpersonal reactions, such as high neuroticism and psychoticism, and high aggression-hostility. In contrast, Dislike of Driving is associated with self-appraisal of attention, with stressful mood states and with daily hassles and life events. This set of correlates suggests that Dislike of Driving is related primarily to the subjective experience of driver stress. The mixture of cognitive and emotional correlates of Dislike of Driving is consistent with the transactional approach to driver stress: the adverse emotions may be caused by the driver’s cognitive appraisal of his/her capacity to deal with the demands of driving.

**Personality dimensions related to accident proneness**

Halakivi et al. (1989) conducted a study to determine which personality dimensions can predict automobile driving performance when selecting safe drivers in a population of young male adults
with little previous driving experience. The results indicated significant personality dimensions which are strongly related to emotionality: involvement in automobile accidents was correlated with uncontrolled behaviour and little respect of social demands, with carefree, adventurous and impulsive behaviour and ignoring of danger signals. The accident-prone drivers were also more easy-going and ready to take a chance. Accident proneness tended to correlate with worrying, poor self-confidence, guilt proneness, repressive feelings, moodiness and hypochondria. However, accident proneness did not correlate with being realistic, practical, steady, concerned but not worried, in contrast to the findings of an earlier study by Cattell and Eber (1964). Thus, the features of personality which predicted motor accidents can be summarised as an insufficiency in the ability to control changes in emotion, emotional lability and excessive dependence on environmental stimuli and psychastenia.

Attributes for personality characteristics which predicted good driving performance were as follows: conscientious, careful, quick to see dangers, not too trustful, self-controlled, persistent, and conscientious, and lacking excessive extroversy. These personality factors were also negatively correlated with involvement in automobile accidents for which the subject was considered responsible.

A.8.3 Control motivation and decision making of drivers

Motivational factors determine what drivers do or what they must do with their skills (Lajunen and Summala, 1995). Among all motivational factors (transient motivation, personality and attitudes towards driving), it is crucial to know which skills are more sensitive to the effects of transient motivational factors (Sivak, 1981). In their study, Delhomme and Meyer (1998) focus on transient motivational factors, extrinsic to driving. Drivers sometimes find themselves at the wheel following a happy or unhappy event related to their work or their social and family life. These factors can give rise to differences in the allocation of attention, changes in the choice of a decision (for example, choosing a more risky alternative), or in the carrying out of a decision (for example, a slower decision). These effects might be understood in the light of motivational consequences of loss of control.

Control motivation and performance

Most individuals spontaneously expect to control the environment by adopting appropriate behaviours (Alloy et al., 1993). A transitory experience of loss of control (for instance, being
unable to resolve a series of logical problems) can have two types of effect on performance in later tasks which have no relation with the initial failure. First, this loss of control can generate an increased motivation to regain control of the environment and result in an improvement in performance. Failure would thus seem to have motivational properties and individuals would seek to increase their freedom by effecting actions on the physical and social environment (Brunstein and Gollwitzer, 1996). The stream of consciousness would seem to be modified in that there is greater concentration on the task (Pennebaker, 1989). Attention, focused on seeking control, would be less able to process peripheral information (Easterbrook, 1959). The information from the environment would be processed in a more detailed and systematic way (Pitman and D'Agostino, 1989) to such an extent, that for certain tasks, the participants thus motivated to gain control, would seem to reach better performances. However, this seeking of control can sometimes lead to illusions about an individual's own control (Friedland et al., 1992). Furthermore, if the regaining of control fails, it results in disruptions in performance, learning and motivation. This triple deficit can be found a long time afterwards in contexts which are different to the initial experience of failure. A repeated experience of non-contingency between actions effected on the environment and reinforcements leads to a state of learned helplessness which, in extreme forms, is similar to depression (Peterson et al., 1993). A common explanation of disruption in cognitive performance following a loss of control is that attentional resources are mobilised on thoughts which have no link with the task (Seiber and Ellis, 1991). Attention could be partly focused on the self and less on the environment, thus limiting the available resources for processing information from the environment (Palfai and Salovey, 1992).

Studies of stress have also focused on cognitive-emotional experience as a response to loss of control (Lazarus, 1993). Under stress, individuals examine alternative options less systematically when faced with a problem and make less optimal decisions (Keinan, 1987). Stress experienced in driving is positively correlated to age, driving experience (Gulian et al., 1990), and to stress experienced in circumstances other than driving. There is a positive link between involvement in an accident and people's level of stress on the day when they had the accident (Isherwood et al., 1982). Individual differences in terms of stress, intrinsic or extrinsic to driving are linked to modifications in information processing, especially in the executive control of attention (Matthews and Desmond, 1995). More generally, emotions, fatigue, and pressure of time would influence the decisions taken in various traffic situations (Näätänen and Summala, 1976).

It is difficult to separate loss of control from concomitant emotional variations (for example, anger, fear, sadness) (Sedek and Kofta, 1990). Extremes of mood (positive or negative) are intrinsically
likely to bring about a disruption in performance. Pressure of time causes similar effects on judgment and decision making with acceleration of cognitive processing, filtration of information and changing strategies (Edland and Svenson, 1993). Generally, time pressure provokes disruption in performance, but sometimes also a greater discriminating ability.

**Integration of control motivation, driving activity and driving experience**

The effect of a change in control motivation on driving activity can be approached from the point of view of information processing related to the road environment (Delhomme and Meyer, 1998). The classic framework of information processing implies a hierarchy of processing levels which range from the most automatised to the most controlled processing (Schneider and Schiffrin, 1977). The multiple components involved in the driving task emphasise the importance of the executive control of attention. In this type of task, an increase in control motivation should have an effect on the higher levels of decision. An increased control motivation, extrinsic or intrinsic to driving, would seem to lead individuals to engage in a more attentive and deliberate exploration of the driving environment. The critical cues of driving situations (presence of other users, and presence of an intersection in particular) should therefore be easier to detect. This increased attention should bring about a more appropriate adjustment to the road environment. Inversely, a decreased control motivation would result in less sustained attention and poorer discrimination of the critical cues of driving situations.

Driving experience develops knowledge schemata which direct attention towards critical cues of the road environment (Delhomme, 1995). Experience enables drivers to make distinctions particularly about aspects of the activity which require dynamic interactions between drivers. Motivational effects related to loss of control should be more of a disadvantage to the less experienced drivers when they have to manage situations which require a heavier cognitive load. Low control motivation should reduce attentional resources towards the environment which would therefore be less available for the processing of complex road information, for example, at an intersection.

In their study, Delhomme and Meyer (1998) found that novices choose to make less speed regulations as a function of critical cues in different driving situations than more experienced drivers do. When choosing speed regulations, novices are also more dependent on their motivational state at that moment. For novices, speed regulation is thus more independent of the characteristics of driving situations. Increased deceleration, when facing dangerous situations in the case of the more experienced drivers, indicates specific self-regulation ability. With even little
driving experience, high control motivation would seem to facilitate the decision to reduce speed. Low motivation to control the environment could be associated with less safe driving. Novices are more impervious to the critical cues of the driving situation and take the context into account less when carrying out speed regulation which leads them to make less safe choices.

The stability of the effect of transient variations in control motivation in more experienced drivers appears to be an important aspect of driving experience. Not only are more complex and appropriate knowledge schemata made available by experience, but also these schemata and patterns are implemented in a way which is less dependent on transient variations in control motivation. An important question is to find out whether the greater independence of the more experienced drivers to the motivational context (as compared with the novices) can be explained solely in terms of automatisation of information processing or whether one needs to consider specific regulation processes of motivational states. Can this not be explained by a particular ability on the part of drivers to regulate their own behaviours (Weinert and Kluwe, 1987)? Perhaps not only high levels of decision (for example, speed choice) but also lower levels of decision (motor skills) could change as a function of control motivation.

Modelling of the driving activity should take more account of drivers’ ability to regulate their own transient motivational states (Delhomme and Meyer, 1998). Unsafe behaviours are often due to a number of different factors, some are due to a lack of driving experience (cognitive and motor-skills) and others are the result of motivational factors (for example, control motivation).

A.8.4 Threat appeals and reckless driving

Due to the fact that fast transportation plays a central role in urban life, driving, particularly among youngsters, is one of the most common risky behaviours which may endanger life. As a result, most of Western societies have invested a lot of material and intellectual resources in planning and developing educational programmes and media campaigns aimed at encouraging youngsters to drive safely. Usually, these efforts attempt to bring into awareness the potential negative consequences of reckless driving (for example, physical injuries and death) via fear-arousal appeals (Ben-Ari et al., 2000).

One of the most relevant issues in applied social psychology research deals with the effectiveness of fear-arousing messages for changing attitudes and behaviours. In their study, Janis and Feshbach (1953) found that a persuasive message, presenting a person with the possibility that he/she was at
risk with regard to the occurrence of an aversive event, was effective in changing his/her behaviour. Leventhal et al. (1965) proposed a model in which the fear aroused by a threatening message leads the person to mentally and behaviourally rehearse recommended actions to avoid the danger and thus reduce the fear. This model has been revised and refined by introducing factors, which are related to the message and the person and that may moderate the effectiveness of threatening messages (Janz and Becker, 1984). McGuire (1985) termed these messages threat appeals and claimed that the severity of the threat may determine its persuasiveness.

Threat appeals have also been used for many years as the main mean to foster careful driving-related attitudes and behaviours (King and Reid, 1990). These appeals usually consist of road-trauma films, showing crashed cars, bereaved parents, or injured people who were involved in car accidents (Dejong and Atkin, 1995). The goal of these appeals is to elicit fear and to remind people of their own vulnerability and finitude while driving in a reckless manner.

However, there is little empirical evidence supporting the effectiveness of driving-related threat appeals. The observed failure of threat appeals to induce safe driving has been explained in a wide variety of terms. Some authors have claimed that defence mechanisms like rationalisation, repression, denial and distancing, may be responsible for this failure (for example, Weilbacher, 1984). Kohn et al. (1982) also proposed that the effectiveness of threat appeals may be moderated by individual factors such as gender and self-esteem. The personality trait of sensation seeking (Zuckerman, 1979a), which has been found to be related to risky driving (Jonah, 1997), may also explain the failure of threat appeals to induce safe driving. A threat appeal may lead people who are high in sensation seeking to perceive driving as a source of thrill, and then, to engage in risky driving.

Ben-Ari et al. (2000) conducted a study to broaden understanding of the motivational basis of reckless driving, and for improving the effectiveness of driving-related interventions. It was found that driving-related threat appeals (for example, a road-trauma film) have a significant impact on both self-reported intentions to drive recklessly and actual behaviour (driving speed) in a driving simulator. Second, these effects were found to have an opposite direction depending on the measurement technique used. Whereas a road-trauma film led to fewer self-reported intentions to drive recklessly than a neutral film, it induced higher driving speed in a simulator. In other words, a threat appeal seems to lead participants to report that they will drive more carefully in hypothetical scenarios, whereas in fact they actually behaved in a less careful way while driving in a simulator. However, all these effects were only found among participants who scored high on the
scale for driving as relevant to self-esteem. That is, the road-trauma film seems to influence reckless driving only when driving is perceived to be relevant to self-esteem. Although behaviour in a driving simulator is a better exemplar than self-report measures of reckless driving, it still does not represent a person’s actual driving behaviour in real life.

When asking people to report on their intentions to drive in hypothetical scenarios, the cognitive system may be focused on analysing and judging the conventional way to behave in these scenarios. That is, people may be mainly concerned with what is expected from them while driving in these scenarios. Therefore, they may be susceptible to the social expectations implied in the film and may be less inclined to report intentions to drive in a reckless way. In contrast, when people actually drove in a simulator, the cognitive system may be more preoccupied with the task of driving and showing off their driving abilities and skills rather than on the potential costs of their reckless driving. Therefore, people may be susceptible to the social demands implied by the road-trauma film and may choose to enhance their self-esteem by driving recklessly. It is worth noting that in both cases, these explanations should be limited to persons who perceive driving as relevant to their self-esteem (Ben-Ari et al., 2000).

The findings can also be interpreted in terms of reactance theory (Brehm, 1966; Brehm and Brehm, 1981). This theory attempts to explain human behaviour in situations involving a threat to perceived freedom. It contends that any reduction or threat of reduction in personal freedom arouses a motivational state, which is directed towards the establishment of the lost or threatened freedom. Generally, reactance arousal is evidenced in behavioural efforts to reassert a threatened freedom, such as rejecting a coercive attempt at attitude change. It is possible that a road-trauma film will be perceived as coercive by the participants, thus reducing the message’s effectiveness and leading to a boomerang effect, increasing behaviours which the film intended to reduce. This explanation, however, is limited to participant’s behaviours in the driving simulator. It cannot explain the effects of the road-trauma film on the self-report questionnaire.

In general, the findings indicate that in certain circumstances, people may make decisions which, although they seem to contradict the biological principle of self-preservation, may serve as symbolic means to protect themselves from the awareness of their finitude. Further studies should examine whether road-trauma films increase the awareness of personal death and the accessibility of death-related thoughts, and compare the psychological effects of threat appeals which involve fear of personal death versus fear of death of others (Ben-Ari et al., 2000).
Though each driver is an unique individual, the traffic setting creates similar forces on all road users. This becomes obvious when comparing people’s biographies or life events. Despite the uniqueness and the individuality of people’s lives, they are composed of similar units like birthdays, growing up, going to school, getting into fights, getting into an accident, getting married, flossing teeth, writing a letter, and so on. One can think of them as various stones to step on, as one wishes or not, in the progression of life. Comparing people’s lives, the similarity and overlap of these bits of life can be seen.

The units of anyone’s traffic life are also cultural elements which all drivers share. Many people fail to recognise this. After all, a driver is encapsulated in his/her driving cocoon made up of his/her thoughts and feelings. These are internal activities, visible to the driver only. The private world of the driver can be lonely and frightening, raked by emotional storms and suspicious thoughts. It is normal to think that one’s thoughts and feelings in traffic are somehow peculiar or unique to oneself. At the very least, one feels one has created them and they feel part of oneself. Yet they are cultural, standardised units of experience shared in common by all drivers.

The mental health of drivers

Self-witnessing reports of one’s private or subjective world as a driver reveals an agitated world replete with extreme emotions and impulses triggered by little acts. Ordinary drivers can display maniacal thoughts, violent feelings, virulent speech and physiological signs of high stress. One driver’s transcript shows the following entry for a particular incident: “My affective behaviour is scared, anxious, fearful, panic stricken, agitated, bothered, irritated, annoyed, angry, mad. I feel like yelling and hitting. My cognitive behaviour is thinking. Oh, no what is he doing? What’s happening? How could he do that? The guy was speeding. My sensorimotor behaviour is that I hear myself saying out loud: Stupid guy! I’m breathing fast, gripping the wheel, perspiring, sitting up straight and slightly forward, my eyes are open and watching straight ahead”.

This incident involved a car cutting into the lane and forcing the car immediately ahead to slam on the brakes causing a chain reaction; however, no collision occurred. Self-witnessing reports of college students collected by James, reveal that each driving trip to campus (averaging less than one hour) is full of incidents of this sort in which near misses occur. Hence, it has become normal and
usual for them to experience stress and panic under everyday driving conditions. A summary of the types of negative reactions frequently mentioned by witnesses includes:

a) Extreme physiological reactions: heart pounding, stopping breathing, muscle spasms, stomach cramps, wet hands, pallor, faintness, trembling, nausea, discoordination, inhibition, visual fixation, facial distortion, back pain, neck cramp.

b) Extreme emotional reactions: outbursts of anger, yelling, aggressive gestures, looking mean and glaring, threatening with dangerous vehicle manipulation, fantasies of violence and revenge against other drivers, panic, incapacitation, distortion, regressive rigid pattern of behaviour, fear, anxiety, delusional talk against non-present drivers and objects.

c) Extreme irrational thought sequences: paranoiac thinking that one is being followed or inspected, talking out loud to other drivers who are not within ear shot, script-writing scenarios involving vengeance and cruelty against ‘guilty’ drivers, denial of reality and defensiveness when a passenger complains of a driver’s error, psychopathic interactions when two drivers alternately tailgate each other dangerously at high speed.

Self-witnessing reports by drivers reveal that driving behaviour is a complex entity occurring simultaneously within three conscious behavioural areas of the individual: affective, cognitive and sensorimotor. Content analysis of the reports shows the driver to be involved in the effort to comply to rules (for example, traffic signs), norms (for example, not following too closely), and roles of driving behaviour (for example, being a bully or a polite driver). In this struggle to comply, three aspects of the driver’s inner world are prominent: compliance in relation to the driver’s feelings (affective compliance), compliance in relation to the driver’s thoughts (cognitive compliance), and compliance in relation to the driver’s sensory and motor responses, such as, sensations, perceptions, motor acts, and overt verbalisations (sensorimotor compliance). These three domains of compliance constitute the driver’s threefold self. Growth, maturity, or expertise as a driver will be a function of the driver’s threefold self.

The driver’s Affective Self

In the following tape-recorded data segment, several dimensions of affect are discernible in the person’s experience during a routine driving episode: “Oh, no, there’s a police car coming up from behind. I hope he did not see me driving fast. Besides, I’m not the only one who is driving fast. If he pulls me over to the side, he has to pull everyone else over too. I’ll be so embarrassed if he pulls me over. Everyone will know that I was breaking the law.”
Contents analysis focuses on the ‘speech act’ value of the components of verbalisations (Searle, 1969; Jakobovits and Gordon, 1974). For instance, “Oh, no” marks an affective stricture or a perception of doom; it indexes an emotional flooding-out. “I hope” marks a religious affection or an idealised picture of reality. “Besides, I’m not the only one” bespeaks guilt and self-justification; it raises the vision of personal catastrophe expressed in “I’ll be so embarrassed…Everyone will know…” Later the same subject displays affections of condemnation or disapproval when another car cuts in front: “Careless and pushy drivers always do things like that”. In another episode, this person expresses anxiety and fear: “I almost sideswiped a car which had been travelling in my blind spot. As I was turning back into the middle lane, I was in a state of mild anxiety. Thinking about what could’ve happened made me scared”. Thus, expressing fear in a driving incident or showing disapproval of another driver are instances of affective driving behaviour.

The driver’s Cognitive Self

Data regarding an individual’s cognitive driving behaviour are obtained from the following entry which was recorded for the same episode: “I should cut down on how fast I’m driving and maintain the required speed limit. I am in the middle lane and yet I am driving like an aggressive person in the left lane. I could be increasing my chance of becoming a victim on the road. If the police pulls me over and gives me a ticket, it’s nobody else’s fault but my own. I should follow the rules. I do not want others to get a bad impression of me and think that I’m a speed demon”.

Reasoning about propriety is evident in “I should maintain the proper speed limit” and “I’m driving like an aggressive person” which also indicates self-evaluation (“aggressive”). Propriety, as well as morality, is involved in the driver’s reasoning regarding the self-attribution of error. (“It’s nobody else’s fault but my own”), and the entry “I do not want others to get a bad impression of me” reveals the person’s image management techniques. In the following entry the driver seems to be overwhelmed with the reasoned consequences of his action: “I am thinking to myself I could have killed the guy back there. I am so careless. He must be swearing at me and saying what an idiot I am. I could have smashed up my brother’s car”.

This self-analysis includes imagining what the others are thinking, feeling, or saying (“He must be swearing…”). Thus, reasoning about a driving situation or attributing an error to oneself are instances of cognitive driving behaviour.
The driver’s Sensorimotor Self

For this segment of the record, the driver spoke of the following in connection with the same episode: “I’ll drive at the required speed limit and get to my destination safely. I am leaning slightly forward in my seat rather than my normal slightly reclined position. I have both hands on the steering wheel rather than my normal one (hand). And I can feel my temperature rising”.

The person is giving some details on motor behaviour and the sensation of getting warmer. Some of this information might be available to an observer or camera (“I am leaning slightly forward in my seat”), but the meaning of this act would remain obscure without the self-witnessing report (“rather than my normal slightly reclined position”) or would require complex instrumentation (“I can feel my temperature rising”). Thus, describing sensations or motor actions are instances of sensorimotor driving behaviour.

The driver’s Threefold Self

There are two perspectives possible and necessary on what people do as drivers, one external, and the other internal. The external view on driving includes road conditions and vehicle manipulation; data on these is obtainable from instruments, measurements and observer evaluation. The internal view on driving is the perspective of the drivers themselves: their sensations, perceptions, verbalisations, thoughts, decisions, emotions and feelings.

In its modern version, behaviourism is committed to a unified theory which tries to deal with external and internal aspects of the self (Staats, 1981; Mischel, 1973). For instance, the concept of personality is defined in terms of built-up repertoires of basic habits. These are actually skills and errors which can be modified through further learning. This acquisition process is going on in three distinct domains of the person: affective, cognitive and sensorimotor (or perceptual-motor). All skills at any level of expertise contain affective, cognitive and sensorimotor features.

One can side-step the dualism issue by defining all human capacities as behaviour. Since perceiving, thinking, and willing or feeling are recognised human capacities, they are defined as behaviour. Similarly, to make a decision after analysing a situation is a behaviour. As well, to feel angry in an incident and desire revenge, is a behaviour.
Since ancient times there has been agreement among philosophers that human capacities are organised into three distinct groups corresponding to the threefold human nature: the will, the understanding and the actions of an individual. Modern psychologists also function within this threefold system of behaviour. What pertains to the behaviour of the will is called affective behaviour and includes affections, feelings, motives, needs and everything which pertains to the goal-directedness of people’s actions. For example, signalling before changing lanes is embedded in an affective context: the driver maintains the motive of avoiding driving errors. In the absence of this motive, errors are committed and the drivers fail to signal. Learning to maintain the motive of avoiding driving errors is an important affective driving skill. Frequently, affective driving errors occur when conflict between motives is experienced, as when a driver is in a hurry and speeds: the feeling of wanting to be cautious and law abiding is weakened by the feeling of urge to hurry and not be too late. The theory of driving behaviour must include the capacity to explain the content and organisation of affective driving skills and errors.

What pertains to the behaviour of the understanding is called cognitive behaviour and includes cognitions, thoughts, reasonings and everything which pertains to the decision-making and analysing aspects of people’s actions. For example, signalling before changing lanes is not only embedded in an affective (motivational) context, but in a cognitive context as well: the driver processes information by common-sense logic. Learning to make correct judgments in routine driving situations is an important cognitive driving skill. Frequently, cognitive driving errors occur when an illogical sequence of interpretation leads to an incorrect decision: “I know there is nobody behind me, therefore I will not bother signalling this time”. This erroneous decision overlooks or ignores several reasons which should be taken into account such as: “There may be somebody in my blind spot, or there may be somebody from the front which might turn in, or there may be a policeman watching”. A comprehensive theory of driving behaviour will have the capacity to identify correct and incorrect decision-making, and specify how cognitions interact with affections to produce overt acts.

What pertains to the individual’s overt actions is called sensorimotor or psychomotor behaviour and includes all experience which is mediated through sensory and motor channels. For example, signalling before changing lanes is a complex psychomotor action involving eye-hand coordination, motor readiness to apply the brakes if needed, twisting of neck to look behind, changes in breathing pattern, and less visible endocrinal and neurological changes. As well, silent or overt verbalisations may occur involving the articulatory system (for example, “Oops, I did not see that car!” or “Ok, now, watch out for that car”). A realistic driving theory should include the
specification of the sequence of psychomotor or sensorimotor actions of drivers and how these are influenced or conditioned by the on-going affective and cognitive context.

One needs to map out the behaviour of drivers under varying social and psychological conditions to arrive at a comprehensive theory of driving behaviour.

Self-witnessing produces a protocol analysis which brings personal awareness. This awareness or new insight provides an opportunity or occasion for self-modification of one’s traffic behaviour from anti-social to peaceful and altruistic. Unwitnessed driving is responsible for traffic insanity and barbaric conditions on public roads. Self-witnessing is the cure for traffic insanity. When drivers verbalise what they are witnessing, as if giving a blow-by-blow description of what’s going on in terms of their feelings, thoughts and actions, then they are maintaining a self-witnessing focus in traffic. The act of verbalising brings awareness of oneself as a driver. Cognitively, it is part of people’s higher mental processes or abilities. Self-awareness produces new knowledge about the self and thus, new reactions. Self-witnessing is a cultural resource since it produces new, more civilised exchanges, mental rituals, and sentiments.
B.1 Unifying Brain Dynamics and Modularity

B.1.1 Overview

How is a human brain functionally organised to achieve adaptive behaviour in a changing world? Grossberg (2000) presents an alternative to the computer analogy, which suggests brains are organised into independent modules, and proposes that brains are organised into parallel processing streams with complementary properties. Hierarchical interactions within each stream and parallel interactions between streams create coherent behavioural representations which overcome the complementary deficiencies of each stream and support unitary conscious experiences. This perspective suggests how brain design reflects the organisation of the physical world with which brains interact. Examples from perception, learning, cognition, and action exhibit the mechanisms by which complementarity is accomplished.

In one simple view, a brain is proposed to possess independent modules, as in a digital computer, and so, for example, humans see by processing perceptual qualities such as form, colour and motion using these independent modules. The organisation of the brain into processing streams (DeYoe and Van Essen, 1988) supports the idea that brain processing is specialised, but it does not imply that these streams contain independent modules. Independent modules should be able to fully compute their particular processes on their own. However, much perceptual data argue against the existence of independent modules, because strong interactions are known to occur between perceptual qualities (Egusa, 1983; Faubert and Von Grunau, 1995; Kanizsa, 1974; Pessoa et al., 1996; Smallman and McKee, 1995). For example, changes in perceived form or colour can cause changes in perceived motion, and vice versa; and changes in perceived brightness can cause changes in perceived depth, and vice versa. Answers to how and why these qualities interact are needed to determine the functional and computational units which govern human behaviour.
Grossberg (2000) reviews evidence that the processing streams in the brain compute complementary properties. The properties of each stream are related to those of a complementary stream much as a lock fits its key, or two pieces of a puzzle fit together. It is also suggested how the mechanisms which enable each stream to compute one set of properties prevent it from computing a complementary set of properties. As a result, each of these streams exhibits complementary strengths and weaknesses. How are these complementary properties synthesised into a consistent behavioural experience? It is proposed that interactions between these processing streams overcome their complementary deficiencies and generate behavioural properties which realise the unity of conscious experiences. In this sense, pairs of complementary streams are the functional units because only through their interactions can key behavioural properties be competently computed. These interactions may be used to explain many of the ways in which perceptual qualities are known to influence each other. Thus, although analogies like a key fitting its lock, or puzzle pieces fitting together, are suggestive, they do not fully capture the dynamism of what complementarity means in the brain. Grossberg suggests that the concept of pairs of complementary processes brings new precision to the idea that both functional specialisation and functional integration occur in the brain.

Why does the brain often need several processing stages to form each processing stream? Accumulating evidence suggests that these stages realise a process of hierarchical resolution of uncertainty. ‘Uncertainty’ here means that computing one set of properties at a given stage can suppress information about a different set of properties at that stage. These uncertainties are proposed to be overcome by using more than one processing stage to form a stream. Overcoming informational uncertainty utilises both hierarchical interactions within the stream and the parallel interactions between streams that overcome their complementary deficiencies. Thus, the computational unit is not a single processing stage, but rather proposed to be an ensemble of processing stages which interact within and between complementary processing streams.

According to this view, the organisation of the brain obeys principles of uncertainty and complementarity, as does the physical world with which brains interact, and of which they form a part. Grossberg suggests that these principles reflect the role of each brain as a self-organising measuring device in the world, and of the world. Appropriate principles of uncertainty and complementarity may better explain the functional organisation of the brain than the simpler view of computationally independent modules. In most cases, evidence for the existence of processing streams and their role in behaviour has been developed by many investigators. The fact that pairs
of these streams exhibit complementary computational properties, and that successive processing stages realise a hierarchical resolution of uncertainty, has only gradually become clear through neural modelling. Through a large number of such modelling studies, it gradually became clear that different pairs of streams realise different combinations of complementary properties. So many streams seem to follow this pattern that it is suggested that complementarity may be a general principle of brain design.

B.1.2 Complementary expectation learning and matching during ‘what’ and ‘where’ processing

Visual processing, from the retina through the inferotemporal and parietal cortices, provides an excellent example of parallel processing streams. Complementary form and motion processing are proposed to be part of a larger design for complementary processing whereby objects in the world are cognitively recognised, spatially localised, and acted upon. The form stream inputs to the inferotemporal cortex, whereas the motion stream inputs to the parietal cortex. Many cognitive neuroscience experiments have supported the hypotheses of Ungerleider and Mishkin (1982, 1983) and of Goodale and Milner (1992) that the inferotemporal cortex and its cortical projections learn to categorise and recognise what objects are in the world, whereas the parietal cortex and its cortical projections learn to determine where they are and how to deal with them by locating them in space, tracking them through time, and directing actions towards them. This design thus separates sensory and cognitive processing from spatial and motor processing.

These hypotheses have not, however, noted that sensory and cognitive learning processes are complementary to spatial and motor learning processes on a mechanistic level. Neural modelling has clarified how sensory and cognitive processes solve a key problem, called the ‘stability-plasticity dilemma’ (Carpenter and Grossberg, 1991; Grossberg, 1999a; Grossberg and Merrill, 1996), and can thus rapidly and stably learn about the world throughout life without catastrophically forgetting previous experiences. In other words, humans remain plastic and open to new experiences without risking the stability of previously learned memories. This type of fast stable learning enables people to become experts at dealing with changing environmental conditions – old knowledge representations can be refined by changing contingencies, and new ones built up, without destroying the old ones due to catastrophic forgetting.

On the other hand, catastrophic forgetting is a good property for spatial and motor learning. People have no need to remember all the spatial and motor representations (notably motor maps and gains) which they used when they were children. In fact, the parameters which controlled their small
childhood limbs would cause major problems if they continued to control their larger and stronger adult limbs. This forgetting property of the motor system should not be confused with the more stable sensory and cognitive representations with which they interact that, for example, help a person to ride a bike after years of disuse.

These distinct 'what' and 'where' memory properties are proposed to follow from complementary mechanisms whereby these systems learn expectations about the world, and match these expectations against world data. To see how people use a sensory or cognitive expectation, suppose they were asked to 'find the yellow ball within one-half second, and you will win a $10,000 prize'. Activating an expectation of 'yellow balls' enables more rapid detection of a yellow ball, and with a more energetic neural response, than not looking for it. Neural correlates of such excitatory priming and gain control have been reported by several laboratories (Motter, 1993; Watanabe et al., 1998; Roelfsema et al., 1998; Kapadia et al., 1995; Hupé et al., 1998; Reynolds et al., 1999; Luck et al., 1997). Sensory and cognitive top-down expectations hereby lead to excitatory matching with confirmatory bottom-up data. On the other hand, mismatch between top-down expectations and bottom-up data can suppress the mismatched part of bottom-up data, and thereby start to focus attention upon the matched, or expected, part of the bottom-up data. This sort of excitatory matching and attentional focusing of bottom-up data with top-down expectations is proposed to generate resonant brain states which support conscious experiences (Carpenter and Grossberg, 1991; Grossberg, 1999a; Grossberg and Merrill, 1996). Paradoxical data about conscious perceptual experiences from several modalities have been explained as emergent properties of such resonant states (Grossberg and Merrill, 1996).

In contrast, a motor expectation represents where a person wants to move, such as to the position where his/her hand can grasp a desired object. Such a motor expectation is matched against where the hand is. After the hand moves to the desired position, no further movement is required, and movement stops. Motor expectations, hereby control inhibitory matching. Inhibitory matching does not lead to brain resonance, so motor processing is not conscious. In summary, sensory and cognitive matching is excitatory, whereas spatial and motor matching is inhibitory. These are complementary properties.

Recent modelling work predicts some of the cells and circuits which are proposed to carry out these complementary types of matching. For example, modelling has suggested how top-down sensory matching is controlled in visual cortex, notably from cortical area V2 to V1, and by extension in other sensory and cognitive neocortical circuits (Grossberg, 1999b; Grossberg and Raizada, 2000).
This top-down circuit is part of a larger model of how bottom-up, top-down and horizontal interactions are organised within the laminar circuits of visual cortex (Figure B.1). The circuit generates top-down outputs from cortical layer 6 of V2 that activate, via a possible polysynaptic pathway, layer 6 of V1. Cells in layer 6 of V1, in turn, activate an on-centre off-surround circuit to layer 4 of V1. The on-centre is predicted to have a modulatory effect on layer 4, because of the balancing of excitatory and inhibitory inputs to layer 4 within the on-centre. The inhibitory signals in the off-surround can suppress unattended visual features. This top-down circuit realises a type of folded feedback, whereby feedback inputs from V2 are folded back into the feedforward flow of information from layer 6 to layer 4 of V1. The modulatory nature of layer 6 to layer 4 connections helps to explain a curious fact about cortical design – despite the fact that the lateral geniculate nucleus (LGN) activates layer 6 of V1 in a bottom-up fashion, a separate, direct excitatory pathway exists from LGN to layer 4 of V1. It is predicted that this direct pathway is needed to enable the LGN to drive layer 4 cells to supra-threshold activity levels, because the indirect LGN-6-4 pathway is modulatory. Neurophysiological, anatomical, and psychological experiments are consistent with these predictions.

Figure B.1 The LAMINART model of bottom-up, top-down and horizontal interactions in the lateral geniculate nucleus (LGN) and visual areas V1 and V2. (Numbers on the right refer to cortical layers) (Grossberg, 2000)
The learning processes which accompany these complementary types of matching are also proposed to exhibit complementary properties. Learning within the sensory and cognitive domain is often match learning. Match learning occurs only if a good enough match occurs between active top-down expectations and bottom-up information. When such an approximate match occurs, previously stored knowledge can be refined. If novel information cannot form a good enough match with the expectations which are read-out by previously learned recognition categories, then a memory search is triggered which leads to selection and learning of a new recognition category, rather than catastrophic forgetting of an old one (Carpenter and Grossberg, 1991; Grossberg, 1999a; Grossberg and Merrill, 1996). In contrast, learning within spatial and motor processes is proposed to be mismatch learning which continuously updates sensory-motor maps (Guenther et al., 1994) or the gains of sensory-motor commands (Fiala et al., 1996; Ito, 1984). Thus both learning and matching within the 'what' and 'where' streams may have complementary properties. As a result, people can stably learn what is happening in a changing world, thereby solving the stability-plasticity dilemma (Carpenter and Grossberg, 1991; Grossberg, 1999a; Grossberg and Merrill, 1996), while adaptively updating their representations of where objects are and how to act upon them using bodies whose parameters change continuously through time (Guenther et al., 1994; Fiala et al., 1996; Ito, 1984).

### B.1.3 More on motor control

#### Vector-integration-to-endpoint model

Viapoint (VP) movements are movements to a desired point which are constrained to pass through an intermediate point. Studies have shown that VP movements possess properties, such as smooth curvature around the VP, which are not explicable by treating VP movements as strict concatenations of simpler point-to-point (PTP) movements. Such properties have led some theorists to propose whole-trajectory optimisation models, which imply that the entire trajectory is precomputed before movement initiation. Bullock et al. (1999) report experiments conducted to systematically compare VP with PTP trajectories. Analyses revealed a statistically significant early directional deviation in VP movements without any associated curvature change. An explanation of this effect is offered by extending the vector-integration-to-endpoint (VITE) model (Figure B.2), which postulates that voluntary movement trajectories emerge as internal gating signals control the integration of continuously computed vector commands based on the evolving, perceptible difference between desired and actual position variables. The model explains the observed
trajectories of VP and PTP movements as emergent properties of a dynamical system which does not precompute entire trajectories before movement initiation. The model includes a working memory and a stage sensitive to time-to-contact information. These cooperate to control serial performance. The structural and functional relationships proposed in the model are consistent with available data on forebrain physiology and anatomy (Cisek et al., 1998).

Figure B.2 The VITE circuit model – thick connections represent the kinematic feedback control aspect of the model, whereas thin connections represent additional compensatory circuitry. (CBM: assumed cerebello-cortical input to IFV stage; c.s.: central sulcus; GO: scaleable gating signal; IFV: inertial force vector; i.p.s.: intraparietal sulcus; OFPV: outflow force and position vector; OPV: outflow position vector; SFV: static force vector; α: alpha motor neuron; γ^d: dynamic gamma motor neuron; γ^s: static gamma motor neuron; Ia: type Ia afferent fibre; II: type II afferent fibre; +: excitation; -: inhibition; ×: multiplicative gating; ť: integration) (Grossberg, 2000)

Inhibitory matching is predicted to occur between a Target Position Vector (TPV) which represents where a person wants to move his/her arm, and a Present Position Vector (PPV) which computes an outflow representation of where the arm is currently. This comparison is proposed to occur at Difference Vector (DV) cells in cortical area 5, which compute how far, and in what direction, the arm is commanded to move. This Difference Vector is, in turn, predicted to be transmitted to cortical area 4, where it is multiplicatively gated by a GO signal which is under volitional control. Turning on the GO signal determines whether the limb will move, and its amplitude scales the speed of movement. The product of DV and GO hereby determined a Desired Velocity Vector
(DVV). Such a DV is predicted to be computed at area 5 phasic cells, and its corresponding DDV at area 4 phasic MT cells. Various other cell types within cortical areas 4 and 5 do not perform inhibitory matching, and might even support resonant states.

The modelling illustrates one way in which ecological analyses of perception and action and computational analyses of neural networks can cooperate to understand control principles which may be embodied by the central nervous system for trajectory formation, and more generally for composing streams of behaviour which are and remain adapted to the perceptible environment. The management of serial performances requires coordination of active internal processes with events arising within a perception action cycle. It is also suggested that time-to-contact (TTC) plays an important role in the control of switching between one element of a planned sequence and a subsequent element.

Learning a TTC threshold for action might be based on noticing signal combinations. This is of interest because the cerebellum, which has long been viewed as a signal-combination detector (Albus, 1971), has recently been implicated experimentally as (also) a site of adaptive timing (Perrett et al., 1993). Work by Bullock et al. (1994) and Fiala et al. (1996) has produced a biochemistry-based model of how phase-advanced action components can be learned by Purkinje cell populations in any predictive task context to which the cerebellum is sensitive. The cerebellar model is based on data from eye-blink conditioning studies, in which the animal must learn how to use specific ambient information to time its response so as to avoid contact of a forthcoming air-puff with its eye. The role of alternative adaptive timing mechanisms should be clarified by future empirical and modelling research aimed at elucidating the changes which occur between early and later stages of practice of sequentially structured actions performed in rich perceptual contexts.

Multiple-paired forward and inverse models

Humans demonstrate a remarkable ability to generate accurate and appropriate motor behaviour under many different and often uncertain environmental conditions. Wolpert and Kawato (1998) propose a modular approach to such motor learning and control to exist within the CNS.

Studies of motor adaptation have suggested that people are able to learn multiple controllers and switch between them based on context. In general, when subjects are brought into a laboratory to undergo a motor learning task in which they must adapt to a visual or dynamic perturbation, they can take many movements to adapt (Welch, 1986). Although the time course of adaptation can
extend over hours, on removal of the perturbation, de-adaptation is often very rapid. In some cases, just removing the subject from the experimental equipment is enough to restore pre-perturbation behaviour. Such asymmetry between learning and ‘unlearning’ suggests that learning may represent adaptation of a new module, whereas de-adaptation represents the switching back to a previously learned stable module.

In agreement with this interpretation is work on re-adaptation. On repeated presentation of a visual (Welch et al., 1993) or dynamic perturbation (Brashers-Krug et al., 1996), subjects adapt increasingly rapidly. This suggests that the retained module for the adapted state is not destroyed by de-adaptation and, moreover, it can be quickly switched on again in response to its introduction.

While the adaptation or switching already described can be attributed to performance errors or knowledge of the consequences of one’s actions, there is evidence that a switching process is also in operation dependent on purely sensory components of the context. Several studies have examined the degree to which two different perturbations can be learned and switched between. In these experiments, one or more perturbations are introduced, and the nature of the perturbation is contingent either on the configuration of the body or on some other sensory cue. For example, when subjects are repeatedly exposed to a prismatic displacement induced by wearing prism goggles, they eventually show adaptive changes cued by the feel of the prism glasses without any prism lenses (Martin et al., 1996). Similarly, context-dependent adaptation can also be seen if cued by gaze direction (Shelhamer et al., 1991), body orientation (Baker et al., 1987), arm configuration (Gandolfo et al., 1996) or an auditory tone (Kravitz and Yaffe, 1972). These studies suggest that subjects can switch immediately between two learned behaviours based on the context.

While the studies described suggest that multiple modules can be learned, they do not address whether two modules can be appropriately activated at the same time. Data for the mixing of two new learned modules based on prism work (Ghahramani and Wolpert, 1997) suggest a specific way that multiple modules are integrated. Using a virtual reality system, a singly visual target location was remapped to two different hand positions depending on the starting location of the movement. Such a perturbation creates a conflict in the visuomotor map which captures the (normally one-to-one) relation between visually perceived and actual hand locations. One way to resolve this conflict is to develop two separate visuomotor maps, each appropriate for one of the two starting locations. A separate mechanism could then combine, based on the starting location of the movement, the outputs of the two visuomotor maps. The internal structure of the system was probed by investigating the generalisation properties in response to novel inputs, which in this case
are the starting locations on which it has not been trained. As predicted by a modular architecture, subjects were able to learn both conflicting mappings, and to interpolate smoothly from one visuomotor map to the other as the starting location was varied. This provides evidence that two modules' outputs can be mixed, as the context is varied between the contexts under which each was learned.

More evidence for the mixing of two new learned modules comes from experiments involving simultaneous activation of alternative expert systems which link (i) picture processing with drawing, and (ii) reading with writing (Adi-Japha and Freeman, 2000). Decentralised modelling predicts both the averaging of action-production times and additive effects of neural noise. These results strengthens the trend to regard control in the brain as centralised at different processing levels, from cognitive systems to motor-control systems and may help with the perennial problem of trying to avoid dividing cognition from perception and action.

![Diagram](https://example.com/diagram.png)

**Figure B.3** Schematic of the multiple-paired forward-inverse model with $n$ paired modules. (dotted lines: training signals; $\times$: signal multiplication) (Wolpert et al., 1998)
Based on the benefits of a modular approach and the experimental evidence for modularity, Wolpert and Kawato (1998) therefore propose that the problem of motor learning and control is best solved using multiple controllers. Each controller consists of a forward model (predictor), inverse model (controller), and a responsibility predictor (Figure B.3). The problem of selecting the appropriate modules is solved by generating a responsibility signal for each module based both on the consequences of performed actions, as estimated by the forward models, and on sensory signals, as estimated by the responsibility predictor. Within each module, the inverse and forward models are tightly coupled, by the responsibility signal, during motor learning. This architecture can simultaneously learn the multiple inverse models necessary for control as well as how to select the inverse models appropriate for a given environment. At any one time, one or a subset of these inverse models will contribute to the final motor command, with the responsibility signals determining the extent to which the output of each inverse model contributes to the final feed-forward motor command.

In principle, multiple-paired forward and inverse models could be located anywhere in the brain. However, many lines of investigation, both theoretical and experimental, suggest that the cerebellum is a very promising candidate (Miall et al., 1993). Wolpert et al. (1998) summarised behavioural, anatomical and physiological data which directly and indirectly support the existence of both inverse models and forward models in the cerebellum. The attraction is that the cerebellum is both quite simple and well documented. It has only one output cell, the inhibitory Purkinje cell (P-cell), and four main classes of interneuron; it is also extremely regular in its cytoarchitecture.

Some imaging studies also indirectly support their hypothesis. For example, mental motor imagery is known to activate the cerebellum (Decety et al., 1990). Both forward models and inverse models are expected to be utilised in mental simulation of the movement. Forward models would be used, in place of the motor apparatus, to simulate the results of non-performed actions on an imaginary controlled object, and inverse models would be required to generate the motor command. In a motor learning study using fMRI, Imamizu et al. (1997) demonstrated activation spots in the lateral posterior part of the cerebellum which persisted after learning to use a new tool was accomplished, even though the motor performance errors had returned to the levels of the baseline scanning periods. Although one cannot tell whether forward or inverse models are learned and reflected in these activation spots, the data certainly suggest that some kind of internal model of an external tool is acquired after motor learning in the cerebellum.
B.1.4 Complementary attentive-learning and orienting-search

Match learning has the great advantage that it leads to stable memories in response to changing environmental conditions. It also has a potentially disastrous disadvantage, however – if one can only learn when there is a good enough match between bottom-up data and learned top-down expectations, then how does one ever learn anything which one does not already know? Some popular learning models, such as back propagation, try to escape this problem by assuming that the brain does only 'supervised learning'. During supervised learning, an explicit correct answer, or teaching signal, is provided in response to every input. This teaching signal forces learning to track the correct answer. Such a model cannot learn if an explicit answer is not provided. It appears, however, that some human and animal learning, especially during the crucial early years of life, takes place in a relatively unsupervised fashion.

Other models do allow 'unsupervised learning' to occur. Here, the key problem to be solved is, that if a teacher is not available to force the selection and learning of a representation which can map onto a correct answer, then the internal dynamics of the model must do so on their own. In order to escape the problem of not being able to learn something which one does not already know, some of these models assume that people do already know (or, more exactly, have internal representations for) everything which people may ever wish to know, and that experience just selects and amplifies these representations (Edelman, 1987). These models depend upon the bottom-up filtering of inputs, and a very large number of internal representations that respond to these filtered inputs, to provide enough memory to represent whatever may happen. Having such a large number of representations leads to a combinatorial explosion, with an implausibly large memory. Thus, although using a very large number of representations can help with the problem of catastrophic forgetting, it creates other, equally serious, problems instead. Other unsupervised learning models shut down learning as time goes on in order to avoid catastrophic forgetting (Kohonen, 1984).

Grossberg (2000) proposes that these problems are averted in the brain through the use of another complementary interaction, which was briefly mentioned before. This complementary interaction helps to balance between processing the familiar and the unfamiliar, the expected and the unexpected. It does so using complementary processes of resonance and reset, which are predicted to subserve properties of attention and memory search, respectively. This interaction enables the brain to discover and stably learn new representations for novel events in an efficient way, without
assuming that representations already exist for as yet unexperienced events. It hereby solves the combinatorial explosion while also solving the stability-plasticity dilemma.

One of these complementary subsystems is just the 'what' stream with its top-down expectations which are matched against bottom-up inputs. When a recognition category activates a top-down expectation which achieves a good enough match with bottom-up data, that match process focuses attention upon those feature clusters in the bottom-up input which are expected (Figure B.4).

Experimental evidence for such matching and attentional processes has been found in neurophysiological data about perception and recognition (Roelfsema et al., 1998; Hupé, 1998; Bullier et al., 1996; Motter, 1994a, b; Reynolds et al., 1995; Sillito et al, 1994). Many behavioural and neural data have been explained by assuming that such top-down feedback processes can lead to resonant brain states which play a key role in dynamically stabilising both developmental and
learning processes (Carpenter and Grossberg, 1991; Grossberg, 1999a; Grossberg and Merrill, 1996; Grossberg 1999b; Grossberg, 1982, 1987; Grossberg et al., 1997).

How does a sufficiently bad mismatch between an active top-down expectation and a bottom-up input drive a memory search, say because the input represents an unfamiliar type of experience? This mismatch within the attentional system is proposed to activate a complementary orienting system, which is sensitive to unexpected and unfamiliar events. Output signals from the orienting system rapidly reset the recognition category which has been reading out the poorly matching top-down expectation (Figure B.4(b) and B.4(c)). The cause of the mismatch is hereby removed, thereby freeing the system to activate a different recognition category (Figure B.4(d)). The reset event hereby triggers memory search, or hypothesis testing, which automatically leads to the selection of a recognition category matching the input better. If no such recognition category exists, say because the bottom-up input represents a truly novel experience, then the search process can automatically activate an as yet uncommitted population of cells, with which to learn about the novel information. This learning process works well under both unsupervised and supervised conditions. Supervision can force a search for new categories which may be culturally determined, and are not based on feature similarity alone. For example, separating the letters E and F into separate recognition categories is culturally determined; they are quite similar based on visual similarity alone. Taken together, the interacting processes of attentive-learning and orienting-search realise a type of error correction through hypothesis testing which can build an ever-growing, self-refining internal model of a changing world.

The complementary attentive-learning and orienting-search subsystems and how they interact have been progressively developed since the 1970’s within Adaptive Resonance Theory, or ART (Carpenter and Grossberg, 1991, Grossberg, 1999a; Grossberg and Merrill, 1996). Neurobiological data have elsewhere been reviewed in support of the ART hypothesis that the attentive-learning system includes such ‘what’ processing regions as inferotemporal cortex and its projections in prefrontal cortex, whereas the orienting-search system includes circuits of the hippocampal system (Grossberg and Merrill, 1996). Data about mismatch cells in the hippocampal system are particularly relevant to this hypothesis (Otto and Eichenbaum, 1992). ART predicts that these interactions between inferotemporal cortex and the hippocampal system during a mismatch event offset the inability of the ‘what’ processing stream to search for and learn appropriate new recognition codes on its own. This deficiency of the ‘what’ stream has been used to predict how hippocampal lesions can lead to symptoms of amnesic memory (Grossberg and Merrill, 1996).
B.1.5 Complementary additive and subtractive intrastream processing

The two types of matching across the ‘what’ and ‘where’ processing streams use different combinations of excitatory and inhibitory neural signals. Complementary processes can also arise within a processing stream. Thus, a processing stream can be broken into complementary substreams. Intrastream complementarity occur within the ‘where’ stream. Here, cortical area MT (middle temporal) activates area MST (middle superior temporal) on the way to parietal cortex. In macaque monkeys, the ventral part of MST helps to track moving visual objects, whereas dorsal MST helps to navigate in the world using global properties of optic flow (Duffy and Wurtz, 1995; Tanaka et al., 1993). These tasks are behaviourally complementary: the former tracks an object moving in the world with respect to an observer, whereas the latter navigates a moving observer with respect to the world. The tasks are also neurophysiologically complementary: neurons in ventral MST compute the relative motion of an object with respect to its background by subtracting background motion from object motion; whereas neurons in dorsal MST compute motions of a wide textured field by adding motion signals over a large visual domain (Tanaka et al., 1993).

Corresponding to the breakdown in area MST into additive and subtractive subregions, area MT of owl monkeys possesses distinct bands and interbands (Born and Tootell, 1992). Band cells have additive receptive fields for visual navigation, whereas interband cells have subtractive receptive fields for computing object-relative motion. Modelling studies have shown how these complementary properties can be used on the one hand, for visual navigation using optical flow information and, on the other hand, for predictive tracking of moving targets using smooth pursuit eye movements (Grossberg et al., 1999; Pack et al, 1999).

Intrastream complementarity is also predicted to occur during sensory-motor control, or ‘how’ processing. To see this, suppose that both eyes fixate an object which can be reached by the arms. Psychophysical (Foley, 1980) and neurophysiological data (Grobstein, 1991; Sakata et al., 1980) suggest that the vergence of the two eyes, as they fixate the object, is used to estimate the radial distance to the object, while the spherical angles which the eyes make relative to the observer’s head estimate the angular position of the object. Distance and angle are mathematically independent properties of the position of an object with respect to an observer. How does the brain compute the distance and angle to an object which the eyes are fixating? A neural model proposes how addition and subtraction can again realise the necessary computations by exploiting the bilateral symmetry of the body (Guenther et al., 1994). In particular, eye movement control pathways give rise to parallel branches, called corollary discharges, which inform other brain systems of the present position of the eyes (Von Helmholtz, 1910-1925). These outflow movement
control pathways have an opponent organisation to control the body’s agonist and antagonist muscles. Neural modelling has mathematically proved that, when both eyes fixate an object, accurate spherical angle and vergence estimates of object position may be derived by adding and subtracting, respectively, the ocular corollary discharges which control the two eyes, while preserving their opponent relationships, at separate populations of cells (Guenther et al., 1994).

These examples illustrate how a rich repertoire of complementary behavioural capabilities can be derived by doing ‘brain arithmetic’, whereby outputs of a processing stage are segregated into additive and subtractive parallel computations at a subsequent processing stage. Such additive and subtractive combinations can occur both between processing streams and within a single processing stream. These simple computations generate very different behavioural properties when applied to different sensory inputs or different stages of a processing stream. The next sections illustrate several ways in which complementary multiplication and division operations may enter the brain’s ‘arithmetic’ repertoire.

### B.1.6 Factorisation of pattern and energy: ratio processing and synchrony

Multiplication and division occur during processes which illustrate the general theme of how the brain achieves factorisation of pattern and energy (Grossberg, 1982). ‘Pattern’ refers to the hypothesis that the brain’s functional units of short-term representation of information, and of long-term learning about this information, are distributed patterns of activation and of synaptic weight, respectively, across a neuronal network. ‘Energy’ refers to the mechanisms whereby pattern processing is turned on and off by activity-dependent modulatory processes.

Why do pattern and energy need to be processed separately? Why cannot a single process do both? One reason is that cell activities can fluctuate within only a narrow dynamic range. Often input amplitudes can vary over a much wider dynamic range. For example, if a large number of input pathways converge on a cell, then the number of active input pathways can vary greatly through time, and with it, the total size of the cell input. Owing to the small dynamic range of the cell, its activity could easily become saturated when a large number of inputs is active. If all the cells got saturated, then their activities could not sensitively represent the relative size, and thus importance, of their respective inputs. One way to prevent this would be to require that each individual input be chosen very small so that the sum of all inputs would not saturate cell activity. However, such small individual inputs could easily be lost in cellular noise. The cell’s small dynamic range could hereby make it insensitive to both small and large inputs as a result of noise and saturation,
respectively, at the lower and upper extremes of the cell’s dynamic range. This noise-saturation
dilemma faces all biological cells, not merely nerve cells. Interactions across a network of cells are
needed to preserve information about the relative sizes of inputs to the cells in the network, and
thereby overcome noise and saturation. This kind of pattern processing sacrifices information
about the absolute amplitude of inputs in order to enable the cells to respond sensitively to their
relative size, over a wide dynamic range. Since the pattern-processing network discards
information about absolute input size, a separate channel is needed to track information about the
total amplitude, or ‘energy’, of the inputs.

Retaining sensitivity to the relative size of inputs can be accomplished by on-centre off-surround
interactions between cells which obey the membrane equations of neurophysiology (Heeger, 1993;
Douglas et al., 1995). In a feedforward on-centre off-surround network, feedforward inputs excite
their target cells while inhibiting more-distant cells. To store inputs temporarily in short-term (or
working) memory, excitatory feedback between nearby cells and inhibitory feedback between
more-distant cells can solve the noise-saturation dilemma. Thus, these networks define mass-
action interactions between short-range cooperative and longer-range competitive inputs or
activities. The mass-action terms of membrane equations introduce multiplication into brain
arithmetic by multiplying cell inputs with cell voltages or activities. Membrane equations respond
to on-centre off-surround interactions by dividing the activity of each cell by a weighted sum of all
the cell inputs (in a feedforward interaction) or activities (in a feedback interaction with which it
interacts. This operation keeps cell activities away from the saturation range by normalising them
while preserving their sensitivity to input ratios.

The ubiquitous nature of the noise-saturation dilemma in all cellular tissues clarifies why such on-
centre off-surround anatomies are found throughout the brain. For example, when ratio processing
and normalisation occur during visual perception, they help to control brightness constancy and
contrast (Grossberg and Todorovic, 1988; Arrington, 1994) as well as perceptual grouping and
attention (Grossberg, 1999b; Grossberg and Raizada, 2000; Gove et al, 1995; Grossberg et al.,
1997). At higher levels of cognitive processing, these mechanisms can provide a neural
explanation of the ‘limited capacity’ of cognitive short-term memory (Grossberg, 1987).

The cooperative-competitive interactions, which preserve cell sensitivity to relative input size, also
bind these cell activities into functional units. Indeed, relative activities need to be computed
synchronously, and early theorems about short-term memory and long-term memory processing
(Grossberg, 1982) predicted an important role for synchronous processing between the interacting
cells. Subsequent neurophysiological experiments have emphasised the functional importance of synchronous brain states (Eckhorn et al., 1988; Gray and Singer, 1989). More recent neural modelling has shown how such synchronised activity patterns can, for example, quantitively explain psychophysical data about temporal order judgments during perceptual grouping within the visual cortex (Grossberg and Grunewald, 1997).

Factorisation of pattern and energy shows itself in many guises. For example, it helps to explain how motor expectations (pattern) interact with volitional speed signals (energy) to generate goal-directed arm movements (Georgopolous et al., 1986; Horak and Anderson, 1984; Bullock et al., 1993), as during the computation of the Desired Velocity Vector (DVV) in the cortical area 4 circuit of Figure B.3. A motor expectation represent where one wants to move, such as to the position where one’s hand can grasp a desired object. Such a Target Position Vector (TPV) can prime a movement, but alone cannot release the movement (Bullock et al., 1998). First the TPV needs to be converted into a Difference Vector (DV), which triggers an overt action only in response to a volitional signal which multiplicatively gates action read-out. The volitional signal for controlling movement speed is called a GO signal. The signal for controlling size is called a GRO signal. Neural models have predicted how such GO and GRO signals might, for example, alter the size and speed of handwritten script without altering its form. As noted, some motor expectations seem to be computed in the parietal and motor cortices. Volitional signals seem to be computed within the basal ganglia. The VITE neural model hereby provides a detailed example of how task-sensitive volitional control of action realises an overall separation into pattern and energy variables.

B.1.7 Cognitive-emotional interactions and attentional blocking

Cognitive-emotional learning enables sensory and cognitive events to acquire emotional and motivational significance. Both classical and instrumental conditioning can be used for this purpose (Pavlov, 1927; Skinner, 1938; Kamin, 1969; Staddon, 1983). For example, during classical conditioning, an irrelevant sensory cue, or conditioned stimulus (CS), is paired with a reinforcing event, or unconditioned stimulus (US). The CS hereby acquires some of the reinforcing properties of the US – it becomes a ‘conditioned reinforcer’ with its own motivational properties. Attentional blocking is one of the key mechanisms whereby animals learn which consequences are causally predicted by their antecedent sensory cues and actions, and which consequences are merely accidental.
During cognitive-emotional learning, at least three types of internal representations interact: sensory and cognitive representations, drive representations, and motor representations (Grossberg and Merrill, 1996; Grossberg, 1982), as depicted in Figure B.5. Sensory representations (S) are thalamocortical representations of external events, like the ones described within the ‘what’ processing stream and include representations of CSs. Drive representations (D) include the hypothalamic and amygdala circuits at which homeostatic and reinforcing cues converge to generate emotional reactions and motivational decisions (Aggleton, 1993; Davis, 1994; LeDoux, 1993). Motor representations (M) include cortical and cerebellar circuits for controlling discrete adaptive responses (Ito, 1984; Thompson, 1988). As noted above, the S representations represent the pattern information in this example. They interact with one another via an on-centre off-surround feedback network which stores their activities in short-term memory, while also solving the noise-saturation dilemma. The D representations supply modulatory energy owing to the action of the following types of learning processes:

a) ‘Conditioned reinforcer learning’ occurs in the S → D pathways, and enables a sensory event, such as a conditioned stimulus CS, to become a conditioned reinforcer which can activate a drive representation D. This may be accomplished by pairing the CS with an unconditioned stimulus US. The CS activates its sensory representation S. The US activates its own sensory representation, which in turn activates the drive representation D. Adaptive weights in the S → D pathway can grow in response to this correlated activity. Future presentations of the CS can hereby lead to activation of D, which controls various emotional and motivational responses.
b) Due to this pairing of CS and US, 'incentive motivational learning' can also occur in the adaptive weights within the D → S pathway. This type of learning allows an activated drive representation D to prime, or modulate, the sensory representations S of all sensory events which have consistently been activated with it in the past. These sensory events are motivationally compatible with D.

c) S → M 'habit learning', or motor learning, trains the sensorimotor maps and gains which control appropriate and accurately calibrated responses to the CS. These processes include circuits such as those summarised in Figure B.3.

Conditioned reinforcer learning and incentive motivational learning combine to control attentional blocking in the following way. As noted above, the sensory representations S are the pattern variables which store sensory and cognitive representations in short-term memory using on-centre off-surround feedback networks. Due to the self-normalising properties of these networks, the total activity which can be stored in short-term memory across the entire network is limited. This is thus, once again, an example of the noise-saturation dilemma. Due to activity normalisation, sufficiently great activation of one sensory representation implies that other sensory representations cannot be stored in short-term memory. In the present example, conditioning of a CS to a US strengthens both S → D conditioned reinforcer and D → S incentive motivational pathways. Thus, when a conditioned reinforcer CS activates its sensory representation S, learned S → D → S positive feedback quickly amplifies the activity of S. This S → D → S feedback pathway supplies the motivational energy which focuses attention upon salient conditioned reinforcers. These amplified sensory representations inhibit the storage of other sensory cues in short-term memory via the lateral inhibition which exists among the sensory representations S. Blocking is hereby explained using incentive motivational 'energy' to amplify conditioned reinforcer CS representations within the self-normalised sensory 'pattern' which is stored in short-term memory. This S → D → S feedback causes a cognitive-emotional resonance to occur. Damasio (1999) has shown how a model of core consciousness can be mapped onto such a cognitive-emotional model ('CogEM' model). In particular, activation of the S ↔ D feedback loop, causing a cognitive-emotional resonance, is predicted to support conscious states.

B.1.8 Rate-invariant speech and language understanding

Factorisation of pattern and energy also seems to play an important role in temporally organised cognitive processes such as speech and language. Here sequences of events are transformed into temporally evolving spatial patterns of activation which are stored within working memories.
(Baddeley, 1986). The ‘pattern’ information stored in working memory represents both the event itself – its so-called item information – and the temporal order in which the events occurred. The ‘energy’ information encodes both the temporal rate and rhythm with which the events occur (Grossberg, 1987). Factorisation of information about item and order from information about rate and rhythm helps humans to understand speech which is spoken at variable rates – rate-invariant representation of speech and language in working memory avoids the need to define multiple representations of the same speech and language utterance at every conceivable rate. This representation can, in turn, be used to learn speech and language codes, or categories, which are themselves not too sensitive to speech rate. Because rate and rhythm information are substantially eliminated from the rate-invariant working memory representation, rate and rhythm need to be computed by a separate process. This is a problem of factorisation, rather than of independent representation, because the speech rate and rhythm which are perceived depend upon the categorical language units, such as syllables and words, which are familiar to the listener. What these language units are, in turn, depends upon how the listener has learned to group together, and categorise, the temporally distributed speech and language features which have previously been stored in the rate-invariant working memory.

Rate-invariant working memories can be designed from specialised versions of the on-centre off-surround feedback networks which are used to solve the noise-saturation dilemma (Grossberg, 1982, 1987; Bradski et al., 1994). In other words, the networks which are used to store spatially distributed feature patterns, without a loss of sensitivity to their identity and relative size, can be specialised to store temporally distributed events, without a loss of sensitivity to their identity and temporal order. The normalisation of these stored activities is the basis for their rate-invariant properties. Thus, Grossberg (2000) predicts that a process like discounting the illuminant, in the spatial domain, uses a variant of the same mechanisms which are used to process rate-invariant speech, in the temporal domain. A key problem concerns how the rate-invariant working memory can maintain the same representation as the speech rate speeds up. It is predicted that the ‘energy’ information which is computed from the speech rate and rhythm can be used to automatically gain-control the processing rate of the working memory to maintain its rate-invariant speech properties (Boardman et al., 1999). In particular, the rate at which the working memory stores individual events needs to keep up with the overall rate at which successive speech sounds are presented. A neural model of this process has been progressively developed to quantitatively simulate psychophysical data concerning the categorisation of variable-rate speech by human subjects (Grossberg et al., 1997; Boardman et al., 1999; Grossberg and Myers, 1999), and to interpret functionally neurophysiological data which are consistent with model properties (Grossberg and
Myers, 1999). In this model, the working memory interacts with a categorisation network via bottom-up and top-down pathways, and conscious speech is a resonant wave which emerges through these interactions.

B.1.9 Summary

Empirical and theoretical evidence support the idea that the brain computes complementary operations within parallel pairs of processing streams, but how these streams are determined and how they interact to generate behaviour is still a topic of active research. The variety of these behavioural processes provides some indication of the generality of this organisational principle in the brain. Interstream interactions are proposed to overcome complementary processing deficiencies within each stream. Hierarchical interactions between several levels of each processing stream are proposed to overcome informational uncertainties which occur at individual processing stages within that stream. Hierarchical intrastream interactions and parallel interstream interactions work together to generate behavioural properties which are free from these uncertainties and complementary insufficiencies. Such complementary processing might occur on multiple scales of brain organisation.

Information, as a technical concept, is well defined for stationary information channels, or channels whose statistical properties tend to persist through time. By contrast, brains self-organise on a relatively fast timescale through development and life-long learning, and do so in response to non-stationary, or rapidly changing, statistical properties of their environments. Grossberg (2000) proposes that hierarchical intrastream interactions and parallel interstream interactions between complementary systems are a manifestation of this capacity for self-controlled and stable self-organisation.
B.2 Current Opinion on Selected Topics in Cognitive Neuroscience

B.2.1 Overview

Cognitive neuroscience has undergone explosive growth. New brain-imaging technologies have allowed researchers to address questions which were in the realm of aimless speculation two decades ago. Further, better computers and new formalisms have led to improved, more detailed models of neural function. It is also becoming increasingly possible to link perception, attention, memory, and other aspects of cognition to neurobiology – which allows the fruits of modern neuroscience to bear on the nature of higher mental function. The study of cognition has moved firmly into the domain of biological science (Squire and Kosslyn, 1998).

The soft warm living substance of the brain and nervous system stands in stark contrast to the rigid metal and plastic hardware of a modern day computer. Not only are the nerve cell units (neurons) self-repairing (to different degrees) and self-wiring under the design built into genes, but they can also promote, amplify, block, inhibit, or attenuate the micro-electric signals which are passed on to them, and through them, and thereby give rise to signalling patterns of myriad complexity between networks of cerebral neurons, and this provides the physical substrate of mind. Such key processes of signalling by one group, or family, of neurons to another is achieved largely by the secretion of tiny quantities of potent chemical substances by neuronal fibre terminals. These neurotransmitters stimulate selected neighbours, with whom they junction, into producing electrical responses which both qualitatively (for example, excitation or inhibition) and quantitatively (for example, by frequency of neurotransmitter release) reflect the patterns of presynaptic stimulation.

In this way, the nerve impulses are passed on from cell to cell. This continuous alteration between electrical and chemical conveyance of signals on their journeys through the pathways of the brain and nervous system, provides a special opportunity for the traffic of electrical impulses to be modulated or blocked as they attempt to jump the gap between one neuron and the next at their junctions, transposed into pulses of chemical substances. A junction between two neurons is, in general, called a synapse. This is the point where selected constellations of neurons from the vast array of neuronal populations can effectively interact, one with another, to filter, edit, integrate, and add precise direction to their interplay of communication (Gregory, 1987).

Neuroscience is contributing to knowledge in virtually all the areas of brain function, varying from neural computation and cortical information processing, the influence of molecular genetics on
brain function, the effects of glucocorticoids released during stress, brain-imaging techniques, and the importance of animal studies, providing insights about even the highest level functions. Numerous studies have already shed light on how the brain develops – from the early steps in the formation of neural connections based on chemical directional cues, to the final steps, guided by neural activity, ensuring that connections are made in an exquisitely precise manner. Furthermore, a major ongoing research effort is focused on the development of speech, hearing, and language abilities in children. Studies emphasize the capacity (which remains even into adulthood) for reorganising the structure and function of cortical maps as the result of injury or of particular kinds of experience. The effects of experience on the brain can be viewed as a continued elaboration of the developmental programme. Another major body of work concerns the organisation of action. Several brain regions participate in motor control and its plasticity, including the parietal cortex, motor cortex, supplementary motor area, cerebellum, and basal ganglia. The basal ganglia appear to be part of a brain system important for learning of action sequences and habits, while the cerebellum is important for motor actions which require well-timed outputs and for perceptual tasks which require precise timing.

Interestingly, recent efforts to integrate psychometric and neurobiological data about personality have stimulated diverse interdisciplinary applications. The dissociation of major brain systems linked to procedural and propositional memory and learning has clarified the clinical distinction between two components of personality: temperament and character. Temperament can be defined in terms of individual differences in percept-based habits and skills (related to procedural memory and learning), which are regulated by the amygdala, hypothalamus, striatum, and other parts of the limbic system. In contrast, character can be defined in terms of individual differences in concept-based goals and values (related to propositional memory and learning), which are encoded by the hippocampal formation and neocortex. Recent descriptive, developmental, genetic, and neurobehavioural studies indicate that at least four dimensions of temperament (harm-avoidance, novelty-seeking, reward-dependence, and persistence) and three dimensions of character (self-directedness, cooperativeness, and self-transcendence) can be uniquely described and functionally dissociated (Cloninger, 1994).
Neuropeptides – neurotransmitters and hormones

The study of peptides occurring in the nervous system, called neuropeptides, is one of the fast growing areas in neurobiology. Novel peptides from the brain with actions related to functions such as pain, analgesia, and sleep are being discovered at an increasing rate. Methods for the identification of peptides in very small amounts, such as are present in the brain, have been developed in recent years. There is also a growing realisation by neuroscientists of the importance of peptides in brain function, based on the ability of neuropeptides to relay messages selectively between particular groups of cells.

Peptides are made of amino acids joined together to form a chain, and since eighteen different amino acids are found in animals, there are many ways of putting these together to make, for example, dipeptides or decapeptides (with ten amino acids). Different peptides can be distinguished because brain cells, and cells elsewhere in the body, have receptors on their surface which can ‘recognise’ a particular peptide. Thus, each peptide used by the brain can carry a particular ‘message’ if it can travel to a nearby or distant site to interact with its receptor. The combination of the peptide with its receptor will induce a change in the cell, for example, excitation or inhibition of the nerve cell. In this way, a peptide can be used by one group of cells to influence or control another group. There are several types of control. A peptide released from the fibre terminal of one neuron into the synaptic cleft to act on the membrane of another neuron is called a neurotransmitter – this type of action is relatively fast and short-lasting. A peptide which is released into the bloodstream and acts on distant cells capable of receiving it, is called a hormone – this type of action tends to be slow and long-lasting. The synaptic action of a neurotransmitter may be modulated by a third party, a neuroregulator or neuromodulator substance, thereby amplifying or attenuating the action of the neurotransmitter (Gregory, 1987).

If one thinks of nerve fibres as like telegraph wires, then hormones are like radio messages. The programme is sent out and can be picked up by anyone who has a radio to receive it. Most hormones are made by endocrine or ductless glands. These glands are controlled by nerve fibres from the central nervous system and also by circulating hormones of other kinds. The brain orders and controls the secretion of hormones, while the brain itself is subjected to the actions of hormones. Furthermore, neurons and their nerve fibres are often named according to the transmitter substance they synthesise and emit at their endings. So, for instance, there are
cholinergic neurons, releasing acetylcholine, and noradrenergic neurons, releasing noradrenaline. Acetylcholine works the muscles of the body, noradrenaline is the transmitter used in the hypothalamus for controlling body temperature, and dopamine is secreted by neurons concerned with waking the brain from sleep.

Hormones secreted by the gonads (sex hormones) are known to exert potent effects on behaviour. In the life of mammals, androgens and their metabolites organise the brain to produce lifelong, irreversible effects on a variety of reproductive and non-reproductive sexually dimorphic behaviours ('organisational' influences) (Goy and McEwen, 1980). Moreover, fluctuations in the levels of these hormones in adulthood, called 'activational' influences, may alter the likelihood that the same behaviours will occur. As in nonhuman animals, human behaviour is subject to activational fluctuations in sex hormones, evident also in cognitive functioning. The most thoroughly investigated influence of activational hormones on cognition has employed the natural variations in estrogen across the menstrual cycle (Hampson, 1990a; Hampson, 1990b). Women performed better on spatial tests in the low-estrogen (menstrual) phase of the cycle than in the high-estrogen (late follicular or midluteal) phase. In contrast, their performance on articulatory-verbal and fine manual skills was better in the high-estrogen phase. Hormone-related variations in cognitive function appear also in men – men have lower testosterone levels in spring than in autumn, and as predicted and found, spatial tasks are performed better by men in spring than in autumn. The susceptibility of cognitive pattern to variations in sex hormones is consistent with the idea that these hormones also play a role in organising the pertinent neurocognitive systems, and thus that some of the sex differences are due to early organising influences. Therefore, it must also be true that some of the variability in the size of cognitive sex differences from study to study is due to variation in the time of day, the season, and the menstrual phase (Kimura, 1996).

The actions of stress on cognition involve multiple mechanisms and different courses. Whereas painful experiences activate opioid pathways (involving endogenous opiate peptides) and emotionally-laden events activate the adrenergic system, the information which the brain records with or without these additional influences appears to involve the hippocampus and temporal lobe, which are highly vulnerable and plastic regions of the brain. Adrenal steroids (hormones), acting via two types of receptors, modulate biphasically the ability of the hippocampal region to store and retrieve information. Under conditions of prolonged stress, these same adrenal steroids, acting in concert with excitatory amino acid neurotransmitters, can, depending on the extent and level of exposure, cause either reversible dendritic alterations or permanent neuronal loss, particularly in the
ageing brain, leading to cognitive impairments involving declarative memory (McEwen and Sapolsky, 1996).

Alzheimer’s disease, the commonest cause of dementia (a severe loss of cognitive function without impairment of consciousness), is ushered in by episodes of forgetfulness about names and faces, and progress within weeks or months to amnesia for recent events and to general disorientation in time and space, and an inability to manage personal affairs. Later stages are characterised by disregard for personal hygiene and nutrition. Brain scans usually reveal severe atrophy of the cerebral cortex which is associated with loss of pyramidal neurons and a shrunken hippocampus. Histological studies also reveal innumerable small amyloid plaques and neurofibrillary tangles. The most striking loss of neurons is in the basal nucleus of Meynert. Up to 90% of the ACh (acetylcholine – a neurotransmitter) neurons are lost from the basal nucleus, as are their projections to the cerebral cortex. Indeed, degenerating ACh terminals seem to contribute to the neurofibrillary tangles. In view of the known relationship between the hippocampal projection and learning, the main thrust of therapeutic research is being directed to the discovery of drugs capable of activating postsynaptic ACh receptors in the temporal lobe – not least because the receptors may survive for much longer than the presynaptic neurons (FitzGerald, 1995).

Development of connections between neurons

In the mature vertebrate nervous system, there is a very large number of neurons and these are extensively interconnected by nerve fibres to form highly complex patterns, which are remarkably constant from one animal to another. These patterns frequently show topographic order, in the sense that one particular set of neurons (for example, the ganglion cells of the retina) ‘project’ their nerve fibres to a second set of neurons (for example, the cells of one of the primary visual centres of the brain), in a fashion which preserves the organisation of the first set of cells in the pattern of the connections which they form with the second set. In this manner the first set of neurons can be said to ‘map’ on to the second. This phenomenon, whereby the neighbourhood relationships of cells in one set are maintained in the connection pattern formed with the other set, is common within the nervous system.

It is known that many of the functions of the nervous system depend on the existence of these orderly mappings. The brain is seen to lay the foundations of even the higher psychological processes before birth. The way the human brain parts grow at first before birth suggests that the interacting nerve cells might make up and coordinate basic rules for object perception, purposeful
movement patterns, and for motive states, without the benefit of experience. In foetal and postnatal stages the brain is increasingly open to selective influences which enter the electrical traffic of the young brain from outside, through the sensory nerves. Chemical communication with the mother through the placenta significantly affects brain development, and subtle refinement of brain structures may also take place in response to gustatory, mechanical, touch, and auditory stimulation. The foetus begins to have some control of this stimulation by moving to change posture, to displace limbs, and to swallow amniotic fluid. One is therefore obliged to pay careful attention to the steps by which nerve cells are linked up into communicating systems, as brain growth is inseparable from mental growth. It is important to note that the facts of brain growth do not imply that mental and physical abilities governed by the brain simply expand and elaborate independently of stimuli, nor, on the other hand, do they give the brain a passive submissiveness to experience (Gregory, 1987).

Ideally, one would like to be able to investigate the development of all the individual constituent nerve fibres which comprise a particular set of orderly interconnecting neurons, together with the micro-environment through which they grow, and the cells they eventually contact. This sort of approach, where the developing system is followed in detail at a single cell level, is currently possible only in certain ‘simple’ invertebrate nervous systems, where small numbers of individually identifiable neurons exist. However, in more complex vertebrate nervous systems a useful approach has been to study classes of neurons which have orderly projections and to try to find general rules which govern the formation of their connection patterns.

The nervous system develops over a certain time, beginning in embryonic life and continuing for a period which depends on the species examined. Within any animal the development of the nervous system is not uniform – different types of neurons develop and form their connections at different times. At any particular time a neuron can only form connections with other neurons which have already developed sufficiently to be able to receive such connections. As a neuron develops, its surrounding environment of glial cells (non-neural supporting cells), other neurons, and axons, will be unique at any given instant, and the role of this environment is crucial. Various mechanisms have been proposed to account for the specific patterns of axonal outgrowth to find their correct target structures, which may be observed in development. For instance, (i) there may occur the expression of particular sets of molecular components in the cell membranes of certain cells along the fibre pathway which are recognised and preferentially responded to by the growing axons, or (ii) fibres may grow preferentially along specific extra-cellular substances in the environment which are manufactured by cells along the appropriate pathway and not, or to a lesser extent, elsewhere, or (iii) there may be passive guidance by certain structural or spatial aspects of the environment which
could lead axons towards the correct target regions, or (iv) the direction of outgrowth of nerve fibres may be influenced by the differential distribution in their local environment of substances produced elsewhere in the system (Gregory, 1987).

The recognition by growth cones of certain cell-surface molecular markers is the basis of the chemo-affinity hypothesis expounded by Sperry in the 1940s. At the simplest level, this idea explains the way in which fibres are guided to their targets along specific pathways by recognizing cellular clues expressed on either glia or other neurons at a few choice points along the pathway followed by the fibres. Once the fibre has reached the general region of its target structure, the recognition of a specific set of cell surface molecules could determine which cell, or cells, the fibre will synapse with. The observations which led to this hypothesis involved the remarkable precision with which the optic nerve fibres of lower vertebrates are able to regenerate, after injury, to their original target sites in the main visual centre of the brain, the optic tectum. In these experiments the optic nerve fibres became disorganised at the point of the lesion and regrew as an abnormally ordered array, yet they nevertheless returned to their original target sites (Gregory, 1987).

Once a growing axon has reached its target structure, how does it know precisely which cell to make contact with? The widely accepted theory has already been outlined in that each axon recognises a specific set of cell-surface molecules on the appropriate target cell. However, there is considerable evidence that nerve connections are not that specific. Indeed, each ingrowing axon will form many synapses with a variety of target cells, each of which will receive contacts from numerous incoming axons. Furthermore, there is good evidence to show that these connections are continually being broken and re-formed, not only as systems develop, but also in what, in the past, has been considered to be the adult, ‘hard-wired’ system. The idea of specific connections must therefore be seen as relative and not absolute. It seems that axons are directed to certain regions of their target structures, producing there a roughly ordered array. When the axons reach the region of their target group of cells, some signal instructs them to stop extending as single axons and to produce multiple branches, each of which bears a mini growth cone. These axon arborisations explore the target region, forming connections with many cells in the appropriate area. This initial exuberant pattern of connections will be topographically ordered, but only at a somewhat coarse level. As development proceeds, the pattern of connections is refined by retraction of the most exuberant branches and the consolidation of synaptic contacts with cells in a very localised region. In this fashion, each axon will form contacts with target cells also contacted by axons which arise from its neighbouring neurons.
The mechanisms which are responsible for this fine-tuning of the initial connection pattern are believed to involve the functioning of the system, commencing with propagating spontaneous activity waves during prenatal development. More changes occur in the cellular structure of the cortex in the first six months after birth than at any other time in development. A large proportion of the millions of axon collaterals spanning the hemispheres and bridging the gap between them are eliminated, segregating intra- and inter-hemispheric association systems. One idea is that synchronous nerve impulse activity may stabilise connections. If an array of neurons is locally stimulated, neighbouring neurons will initiate impulses at the same time, while other cells in the array will be silent. If several of the stimulated neurons contact the same target cell, then it will receive an increased number of simultaneous impulses and is highly likely to respond. In contrast, a target cell which has only few contacts from the stimulated neurons, will remain silent. In time, contacts which rarely cause their target cell to respond may become redundant to the function of the system and eventually be withdrawn, so that each neuron then concentrates its contacts on those cells where it achieves maximal effect. Such a mechanism could gradually consolidate the contacts made by neighbouring neurons on neighbouring target cells. It is now clear that the anatomy and function of human cerebral cortices are very variable. People differ in the pattern of their mental abilities because their brains grow in different forms. Some of this diversity of human minds will be pre-programmed in a great variety of outcomes of gene expression in nerve tissue development, but the same brain processes are also influenced by foetal and postnatal environments (Gregory, 1987).

Reaction time, neural codes and cortical organisation

Like all other animals, humans can only ‘experience’ the immediate past. Many scores of milliseconds must pass before any change in the world can be registered by a sense organ or interpreted by the brain. This perpetual lag behind the world, measured from the moments at which changes actually occur and the moments at which people can apprehend them, has become known as reaction time. Like all other successful animals, humans overcome their temporal lag behind events in the external world through priming or learning to predict what will happen next. When people can do this they can have reaction times of zero milliseconds, responding to events just as soon as they occur, or they can even take leaps into the future, gaining an edge over a rapidly changing environment, or over rapidly moving adversaries, by anticipating events which have not yet occurred.
Apart from the speed with which electrical impulses travel along nerve fibres, another factor limiting human reaction time is the presence of membrane time constants at synapses. Each neuroreceptor channel in the post-synaptic membrane may be open for only a brief period (for example, 1 microsec) as the neurotransmitter rapidly dissociates and is inactivated, or may remain open for much longer periods (for example, 1 sec) depending on the ion channel (Na\(^+\), K\(^+\), or Cl\(^-\)) concerned. Cortical neurons exhibit time constants of 8-20 msec (Shadlen and Newsome, 1994). The movement of ions through the ‘hole’ in a neuroreceptor protein molecule (once bound to a specific neurotransmitter) either into or out of the interior of the post-synaptic cell generates excitatory or inhibitory synaptic potentials and, from this pattern of impingement of electrical signals (information) onto its dendrites and cell body (inhibitory inputs), the target neuron will be triggered to fire its own action potential, or remain quiescent, as appropriate according to the intensity of the excitatory and inhibitory signals received. It is important to note that reaction times do not provide measurements of the time necessary for sets of nerve impulses generated in the sense organs to activate those parts of the brain that, in turn activate the muscles. They rather measure the duration of operation of processes of active, predictive control, by means of which people organise responses which anticipate, and pre-empt, fast changes in the world (Gregory, 1987).

Although it is generally agreed that neurons signal information through sequences of action potentials, the neural code by which information is transferred through the cortex remains elusive. In the cortex, the timing of successive action potentials is highly irregular, and the interpretation of this irregularity has led to two divergent views of cortical organisation. On the one hand, the irregularity might arise from stochastic forces (arising from processes not yet understood). If so, the irregular interspike interval (ISI) reflects a random process and implies that an instantaneous estimate of spike rate can only emerge from the pooled responses of many individual neurons. In keeping with this theory, one would expect that the temporal pattern of spikes conveys little information. Alternatively, the irregular ISI may result from precise coincidences of pre-synaptic events. In this scenario, it is postulated that the timing of spikes, their intervals and patterns can convey information. According to this view, the irregularity of the ISI reflects a rich bandwidth for information transfer (Shadlen and Newsome, 1994).

Understanding of cortical organisation and interpretation of neurophysiological data depend critically on whether neurons convey a noisy rate code or a precise temporal code. Is it reasonable to expect the average discharge rate of a neuron in the visual cortex to convey information about a visual stimulus, or should one attend to particular patterns of spikes? If the variable ISI reflects a precise temporal code which must be propagated through the cortex, the pattern of cortical
connectivity should emphasise divergence, and redundancy should be avoided. Moving downstream in a cortical pathway, therefore, one would expect fewer neurons to covary their responses under similar stimulus conditions, a view that probably demands reduction, if not outright elimination, of redundancy in the form of columnar organisation. In fact, one would expect to see progressively less spike rate modulation at all, as rate modulation can only muddle a temporal code with spurious coincidences (Abeles et al., 1993; Abeles and Gerstein, 1988). Alternatively, if synaptic integration produces a truly random ISI, the neural code consists simply of modulations in spike rate. As any one neuron provides a poor estimate of the instantaneous spike rate, the cortex must use ensembles of neurons to represent the same information. This view demands a reiterated organisation of redundant, column-like modules, even in higher cortical areas.

Clustering of neurons with similar response properties (redundancy) is a well-established principle in primary sensory and motor areas of the cortex, and is beginning to receive attention from investigators working on higher cortical areas as well. By analysing patterns of connectivity revealed by local biocytin injections, Amir et al. (1993) found a patchy organisation of horizontal connections reiterated in striate, extrastriate and parietal cortex of the macaque monkey. Similar observations have been made in inferotemporal (IT) (Saleem et al., 1993), frontal and limbic cortex (Goldman and Nauta, 1977), suggesting common organisational principles that are consistent with a redundant coding strategy (Lund et al., 1990). Recent physiological data from IT cortex also support this point of view; nearby neurons, probably organised in the form of columns, appear to share preferences for similar features of visual objects (Fujita et al., 1992; Tanaka, 1993) and faces (Perrett et al., 1984; but see Gawne and Richmond, 1993). Simple statistical considerations reveal that the instantaneous rate from an ensemble of 100 or so neurons can be estimated reliably within a single ISI. Shadlen and Newsome (1994) suggest that neuronal pools of this size may comprise the fundamental signalling units of cerebral cortex forming cylindrical columns.

However, the acid test for any theory of the neural code is to establish a connection to behaviour. In the cortical areas MT and MST, fluctuations in spike rate have been shown to correlate with an animal’s decisions in a motion discrimination task. Examples abound in the motor cortex for a connection between neural discharge rate and behaviour. Furthermore, some neural structures convey information in the timing of successive spikes. The best examples are probably from brainstem auditory pathways, where spikes may be time-locked to peripheral events. However, in the cortex, as mentioned previously, many inputs affect a neuron, and a single pre-synaptic spike has little bearing on the exact timing of a post-synaptic spike. But, this is not true if certain inputs
have privileged contacts. For example, a single spike from the thalamus may induce a time-locked spike in visual cortex consistent with mono-synaptic excitation (Shadlen and Newsome, 1994).

Experiments carried out on the visual cortex of awake-behaving monkeys suggest that dynamic grouping and the attentional control of salience may be closely related. Cells within clusters responding to the attended stimulus synchronised their activity to a greater extent than did cells within clusters responding to unattended stimuli. This increased synchronisation was seen for high-frequency activity in the gamma-band (35-90 Hz). An increase in high-frequency synchronisation would be a good way to increase the impact of the population activity of a cluster, because the rapid decay of post-synaptic potentials ensure that inputs will summate at their projection sites much more effectively if they are synchronised. It has also been proposed as a possible neural correlate of consciousness. If one is more conscious of attended activity then evidence that attended activity is more synchronised suggests that one is more conscious of synchronised activity (Philips, 2001).

Much of neural network research builds on the basis that neurons are simple linear threshold units, completely neglecting the highly dynamic and complex nature of synapses, dendrites and voltage-dependent ionic currents (Koch, 1998). How do single nerve cells multiply, integrate or delay synaptic inputs? What is the repertoire of computational operations available to individual nerve cells? How accurately can information be encoded in the voltage across the membrane, in the intracellular Ca\(^{2+}\) concentration or in the timing of individual spikes?

Additional topics which neural network researchers should attend to earnestly, include the linear cable equation, passive dendritic trees and dendritic spines, chemical and electrical synapses and how to treat them from a computational point of view, non-linear interactions in passive and active dendritic trees, the Hodgkin-Huxley model of action potential generation and propagation, phase space analysis, linking stochastic ionic channels to membrane dependent currents, Ca\(^{2+}\) and K\(^{+}\) currents and their role in information processing, the role of diffusion, buffering and binding of calcium and other messenger systems in information processing and storage, short- and long-term models of synaptic plasticity, simplified models of single cells, stochastic aspects of neuronal firing, the nature of neural code and unconventional models of computation involving molecules, puffs of gas, or neuropeptides (Koch, 1998). For example, sub-groups of the more than 50 known neurotransmitters produce different categories of effect in both qualitative and quantitative respects, and many co-exist in different neuronal populations, and can be seen to provide a chorus of informational voices, each adding to the final output of the brain and nervous system.
Furthermore, apart from the longer established post-synaptic neuroreceptors, the existence of pre-synaptic neuroreceptors, being primarily concerned with controlling the extent of neurotransmitter release, add yet another dimension to the theory of neural code.

The possibility exists that the neuron is a multi-channel device, a cable rather than a wire. Since nerve impulses are relatively slow moving, each successive impulse might be rich in information. A multi-channel neuron has the power to convey, with each single all-or-nothing impulse, graded information. For example, to 20 discrete channels, one can assign 20 distinct tiers of meaning, and each channel can thus ‘mean’ a level of intensity between 1 and 20. The phenomenon can easily escape detection because such a neuron appears, to conventional instruments, to convey only the classically blank, binary impulse that is so confidently presented on the first page of every neurobiology textbook. One can visualise many parallel tracks – possibly linear, possibly helical. Linked receptors are commonplace. The molecular structure of the K+ channel has been published recently, and so researchers are now finally working at the level where a multi-channel membrane can be detected. It is a theoretical construct but if each single impulse carries information, then the computational burden on the nervous system is vastly reduced, and the physiological meaning of intensively studied structures like the synapse suddenly changes (Koch, 1998, Rieke et al., 1999).

Koch also discusses several speculations for non-neural computation in the brain, ranging from molecular computing below the level of a single neuron to the effects of chemical diffusants (for example, nitric oxide, calcium ions, carbon monoxide) on large numbers of neurons. Although it appears that the neuroscience community has neglected this entire area, Koch points out that there are no good reasons for doing so.

**Transgenic approaches to brain function**

The modification of behaviours from the simplest reflex responses in invertebrates to the highest-order cognitive processes in humans is thought to involve changes in neurons at the molecular level (Hawkins et al., 1993). Until recently, investigation of the molecular processes which control behaviour in the mammalian brain has been difficult. However, this has changed since a number of methods have emerged by which the mouse genome can be altered in a directed or ‘reverse’ genetic manner (Grant and Silva, 1994; Takahashi et al., 1994; Joyner, 1993). Reverse genetics refers to a set of techniques, such as transgenesis and gene targeting, in which a single cloned gene is used to generate a line of mice with an alteration specifically in that gene. The reverse genetic approach offers a number of advantages in the molecular study of cognition.
First, as only a single gene is altered, this approach offers the highest degree of molecular specificity. One can inactivate distinct subtypes of receptors or cell signalling molecules like protein kinases, for which specific pharmacological blocking agents may not be available. Second, reverse genetics allows subtle aspects of molecular function to be modulated. For example, single point mutations can be introduced that alter the ion flux of a channel or activation properties of a protein kinase. These in turn may produce more varied and subtle changes in neuronal physiology than can be obtained by other methods. Finally, once created, a mutant mouse strain provides an unlimited supply of animals with the identical molecular lesion for electrophysiological, anatomical and behavioural studies. Genetically modified animals thus provide an effective means for investigating the relationship between different aspects of neuronal function and cognitive processes, such as learning and memory (Mayford et al., 1995).

Although the study of mutant mice is beginning to provide insights into many aspects of the molecular basis of cognition, the application of these methods to mental processes is only beginning. The power of genetics is its ability to dissect the sequence of steps involved in a particular physiological pathway. This approach should prove helpful in defining the various parallel and distributed pathways of processing by eliminating not only key molecules, but also key functional pathways. The ability to genetically disrupt these higher-order neuronal functions may in turn lead to the development of animal models for some of the major neurological and psychiatric disorders. The rapidly developing technology for manipulation of the mouse genome holds the possibility of modifying specific regions of the mammalian brain at specific moments in the lifetime of the animal. In this way, one may begin to bridge the gap in understanding how the genes control neuronal function and how neuronal function, in turn, controls behaviour (Mayford et al., 1995).

B.2.3 Aspects of brain development and structure

Functional and structural brain mapping

Perception, action, cognition and emotion can now be mapped in the brain by a growing family of techniques. Positron emission tomography (PET), functional magnetic resonance imaging (fMRI), event-related electrical potentials (ERP), event-related magnetic fields (ERF) and other non-invasive imaging techniques are rapidly evolving and providing an increasingly rich literature on the functional organisation of the human brain. Spatially, temporally, physiologically and cognitively accurate computational models of the neural systems of human behaviour are the
ultimate objective of functional brain mapping. This objective will be reached only through integrating the diversity of modern brain-mapping methods (Fox and Woldorff, 1994).

The complementarity of imaging modalities makes within-subject integration of brain maps scientifically and clinically appealing. Simple fusion (co-registration) of tomographic modalities (for example, PET and MRI) is relatively straightforward. The fusion of activation maps with three-dimensional renderings of brain anatomy is still more useful. These integrated maps illustrate the variability of the cortical folding patterns and the degree to which they can predict functional location. Although graphically very appealing, the shortcoming of such fusions is their inability to quantify the variability they so elegantly illustrate. Without some means of anatomical quantification, within-subject mergers of functional and structural maps remain essentially pictorial. Fusion of surface detection methods (for example, ERP and ERF) with volumetrically true techniques (for example, PET, MRI, and CT: X-ray computed tomography) is far more challenging than mergers of two tomographic modalities. The problem becomes tractable only through modelling (Fox and Woldorff, 1994).

A long-term goal of human-brain mapping is the creation of physiologically and anatomically accurate spatial and temporal models of the neural systems underlying human behaviour, if deemed useful for a particular application. To accommodate the multiplicity of neural systems, as well as population variables (such as gender, handedness and native language), multiple models will be needed. Population models derived from brain images will necessarily be composites, formed from groups of subjects. Integration of images from different subjects is based on anatomical normalisation. Anatomical normalisation, in turn, is based upon the use of one or more standard ‘anatomical spaces’, within which data can be integrated. Modelling the neural systems of cognition from brain-imaging studies is another challenge. Giving rapid access to the cumulative knowledge of the field, databases should be a powerful resource. Spatial and temporal models are needed to express the neural activation patterns associated with specific behaviours (Fox and Woldorff, 1994).

Brain development in infancy

Since the pioneering work of the 1950s, experimental studies of infants’ perceptual and cognitive abilities have played a crucial role in nature-nurture debates within psychology. Over the past decade or so, the field has largely moved beyond this dichotomy, and the focus of current research has shifted to using infancy data to analyse the earliest precursors of later cognitive abilities.
For example, with regard to face-recognition abilities, the relevant question is no longer whether the ability to recognise a face is ‘innate’, but rather how do the primitive abilities of the newborn shape and guide the subsequent development of specialised face-recognition processes. Viewed from this new perspective, infancy research is beginning to take on a central role in understanding the emergence of a variety of specialised cognitive abilities observed in the adult.

Until recently, the majority of researchers studying the perceptual and cognitive development of infants did so largely without taking account of underlying brain growth. Indeed, some believed that such considerations were distracting or even irrelevant. However, a number of recent technological and theoretical developments encourage the view that the time is now ripe for exploring the neural basis of cognitive development in infants. Furthermore, the interchange of information between cognitive development and developmental neurobiology may be essential for the advancement of both fields (Johnson, 1994).

New technological and methodological developments include the increasing availability of non-invasive functional brain-imaging technologies mentioned earlier, and the increasing use of the ‘marker task’ method. This latter method involves the use of specific behavioural tasks which have been consistently linked to one or more brain region(s) or pathway(s) in adult primates (including humans) by neurophysiological, neuropsychological and/or brain imaging studies (Johnson, 1998). By studying performance on such tasks at different ages and in different contexts, the researcher can gather evidence regarding the development of functional brain systems.

Despite recent advances in the cognitive psychology and cognitive neuroscience approaches to infant development, the importance of a developmental approach to cognitive neuroscience in general should not be underestimated (Johnson, 1993). The adult human brain is composed of a complex series of hierarchical and parallel systems which have sometimes proved difficult to analyse. A developmental cognitive neuroscience approach offers not only the opportunity to study a simpler, more primitive, version of the human brain-mind, but also a glimpse of the increasing complexity of the cognitive system as new brain pathways and structures are added to earlier-developing ones.

For instance, in the case of vision and visual orienting, immaturities in peripheral sensory systems (such as the eye and optic tract) clearly place limits on the perceptual capacities of infants. Nevertheless, the issue remains whether most of the limitations on perception in young infants can
be accounted for in terms of these peripheral factors, or whether maturation of the central nervous system is the primary limiting factor. A number of models, in which infants' preferences between abstract visual patterns are predicted by hypothesised limitations on acuity resulting from peripheral receptors, have been proposed. However, observed differences in spatial and chromatic vision between adults and infants turn out to be significantly greater than predicted from these models, indicating that neural immaturity is a significant contributing factor (Banks and Shannon, 1993).

Given that at least some of the advances in infants' visual capacities can be attributed to developments within the central nervous system, the next question is exactly which brain structures are involved. Evidence from neuroanatomical studies of postmortem tissue (Conel, 1939-1967) and positron emission tomography (PET) imaging (Chugani et al., 1987) indicate that the human cerebral cortex undergoes considerable postnatal development. Several authors have attributed changes in perceptual and visual-orienting abilities over the first few months of life to a general shift from subcortical to cortical visual information processing, and more recently to the sequential development of different cortical pathways (Johnson, 1994). Even infants lacking an entire cerebral hemisphere (as a result of surgery for epilepsy) will readily make saccades to conspicuous targets in the half-field contralateral to their lesion, although these saccades tend to be a little longer in latency than normal. This result indicates that subcortical circuits can support orienting in the absence of the cortex, and lends some credibility to the view that visual orienting in the newborn could be supported largely by subcortical (collicular) circuits.

Shifts of visual attention can occur independently of eye and head movements, so-called 'covert' visual attention. One paradigm for studying such shifts of attention in adults involves a briefly flashed cue in a peripheral spatial location, resulting in facilitation of detection and responses to targets at that location. In infant marker task versions of this procedure, four-month-old infants show both facilitation and subsequent inhibition of return to a cued location, whereas two-month-olds do not. Interestingly, four months of age corresponds to the period of most rapid postnatal development of the parietal cortex (Chugani et al., 1987), suggesting that the maturation of this region of cortex may be an essential prerequisite for covert shifts of attention (Johnson, 1994).

One of the most striking observations that Piaget (1954) made, is that if an object is hidden from view, then it seems to cease to exist for infants under about eight months of age. He used the successful retrieval of a hidden object by older infants as evidence that they had developed a concept of object permanence, in that objects continue to exist even when hidden from view. However, infants as young as three and a half months are surprised by impossible events involving
object movement. One possible explanation for the pattern of data observed is that a stronger
representation of an object is needed in order to drive more complex behaviours, such as those
required for successful performance in Piaget’s original object-permanence task. In tasks which
relatively simple motor commands are required, such as shifts of eye gaze, evidence for object
permanence is seen earlier in development.

Another domain of infant development which has benefited from a cognitive neuroscience approach
is memory. The phenomenon of infantile amnesia has led to speculation that the brain mechanisms
necessary for the long-term storage of information, most probably in the limbic system, are not fully
functional for the first year or two of life. Recent evidence from both cognitive and neuroscience
studies have cast some doubt on this previously commonly held view. Rather, the protracted
maturation of a limbic-dependent recognition memory system has been proposed to account for
infantile amnesia. This proposal is consistent with evidence from event-related potential studies
showing that visual recognition memory is relatively slow to develop in human infants (Nelson and
Deregner, 1992). However, another recent study has provided evidence for limbic-dependent
recognition memory in infant monkeys in the first month of life, and shows that this ability is
impaired by concurrent damage to the amygdaloid complex and hippocampal formation
(Bachevalier et al., 1993). These results again suggest that limbic structures make a significant
contribution to visual recognition memory even at this very early age, and that other neurocognitive
systems that mature more slowly may be responsible for the previous findings.

Given that currently no strong evidence exists for a change in the mechanisms of memory between
infants and adults, how is one to account for infantile amnesia? Rovee-Collier (1993) proposes that
infantile amnesia can be accounted for by two attributes of infant memory – its extreme sensitivity
to context (such as the colour of the walls around a testing apparatus), and its susceptibility to
updating by experiences that occur shortly after the event (Boller and Rovee-Collier, 1992).
Clearly, further research on infant memory and its neural basis is required to test these hypotheses.

To conclude, infants are born with a number of primitive tendencies and sensory limitations which
constrain the possible developmental paths that can be taken by powerful perceptual and cognitive
learning systems. Further, because only parts of the adult cognitive system are in place during
early infancy, and certain brain pathways and structures are slower than others to develop, the study
of infants offers a unique opportunity to unravel how complex hierarchical and parallel competing
brain pathways give rise to adult cognition. The study of infancy is thus central to significant
further progress in the neurosciences (Johnson, 1994).
Insights from instinctive behaviour of animals

Most scientists recognise that all behaviour is influenced both by the animal’s genetic make-up and by the environmental conditions which exist during development. However, the extent to which the influences of nature and nurture determine behaviour varies greatly from activity to activity and species to species (Gregory, 1987). For example, the vocalisations of pigeons and doves are relatively stereotyped and characteristic of each species, and are not influenced by auditory experience after hatching. The vocalisations of other birds, however, may depend upon such experience, as in the strongly imitative birds, or they may be partly influenced by experience. For example, chaffinches will learn the song they hear during a particular sensitive period of early life, provided it is similar to the normal song. While the influence of particular genes may be necessary for the development of a behaviour pattern, it is never a sufficient condition. All types of behaviour require a suitable embryonic environment for the correct nervous connections to develop. Normally, the physiological medium provided by the parent is designed to ensure that normal development of both embryo and juvenile occurs. Thus, a chaffinch is normally reared in an environment in which it inevitably hears the song of other chaffinches, and so it develops the song that is characteristic of its own species.

Even apparently stereotyped activities may, upon closer examination, be shown to be influenced by the environment. For example, the newly hatched chicks of herring gulls peck at the tip of the parent’s bill, which bears a characteristic red spot on a yellow background. The chick’s behaviour induces the parent to regurgitate food. The behaviour is typical of all newly hatched chicks, is performed in an apparently stereotyped manner, and would appear to be a classic example of instinctive behaviour. Upon closer examination, however, it can be seen that the initial behaviour of individual chicks varies considerably in force and rapidity of pecking, angle of approach, and accuracy. As the chicks gain experience their pecking accuracy improves, and the pecking movements become more stereotyped. Some of these changes are due to maturation. The chicks become more stable on their feet as their muscles develop, and their pecking co-ordination improves. Some of the changes are due to learning. Initially the chicks peck at any elongated object of a suitable size. Although the red spot on the parent’s bill is attractive to them, it is not their only target. Once the chicks begin to receive food they learn to exercise greater discrimination. It is not surprising that the behaviour of different chicks develops along similar lines, because in the natural environment they are all confronted with a similar situation.
and experience in similar situations lead to similar results, and the behaviour of the older chick consequently becomes more and more like that of its peers.

The concept of instinct has undergone many changes over the years. Whereas, at one time, instinctive behaviour was seen as inborn, stereotyped, and driven from within, the modern approach is to treat the innate, the reflex, and the motivational aspects as separate issues. While much animal and human behaviour is innate in the sense that it inevitably appears as part of the repertoire under natural conditions, this does not mean that genetic factors are solely responsible. Maturational factors and modes of learning which are characteristic of the species may be just as important. Much of the nature-nurture controversy results from a failure to recognise the vast complexity of developmental processes (Gregory, 1987).

B.2.4 Learning and representation in speech and language

Speech and other central auditory processes

One of the primary objectives of the modern field of cognitive neuroscience is to understand the relationship between basic neurobiology and higher cognitive functions. One of the fundamental assumptions of this field is that through investigating the structure/function relationship for higher cortical processes, researchers can further the understanding of both neurobiology and cognition. That is, cognitive science can benefit from the application of a physiological framework to assess the plausibility of theoretical models of cognition. Similarly, insights provided by considering the functional role of the brain and its constituent parts will guide and constrain neurobiological models.

Recent advances in neuroimaging, which allow for the real-time observation of brain activity, suggest that many higher cognitive processes can be characterised by the salient perceptual features and learnt associations that subserve them. In addition, behavioural and neuropathological studies in humans and animals provide insights into the existence of processing subsystems that may relate to the functional organisation of the brain. Miller et al. (1995) focus on these issues, with a particular focus on the neuroanatomical localisation of speech processes, the role of rapid acoustic changes in speech, and the use of animal models for investigating the neurobiological substrates of speech and language.
Independent of their value in clinical settings, recent evidence from neuroimaging studies have also provided new and intriguing information concerning the neurobiological processes subserving speech and language (Binder et al., 1994; Zatorre et al., 1992; Fiez, 1994). Results from these studies suggest that areas of the brain that have typically been considered to be specialised for speech and language, may instead become active in response to more basic processing demands which may subserve both linguistic and non-linguistic processes. For example, in an attempt to identify the extent to which the superior temporal gyrus of the temporal lobe was specific to the processing of semantic information, fMRI recordings were made while subjects listened to noise, pseudowords, words and fluent speech (Binder et al., 1994). Contrary to expectations, results from this study showed that activation of the superior temporal gyrus occurred for both the noise and speech stimuli. In addition, although the activation of the superior temporal gyrus was significantly higher for the speech versus noise conditions, no consistent semantic-specific differences occurred in response to the different speech stimuli (pseudowords, words and fluent speech). It is suggested that these results emphasise the role of the superior temporal gyrus in the perception of acoustic-phonetic features of speech, rather than processing of semantic features (Binder et al., 1994).

The aforementioned PET and MEG data from normal adults are consistent with the results of neuroimaging studies conducted on individuals with speech, language and reading disorders. In numerous studies it has been convincingly argued that children and adults with language-based learning disabilities (L-LDs) show specific deficits in processing rapidly changing auditory and visual information (Farmer et al., 1995). Based on MRI studies, it has been found that L-LD individuals show a lack of the expected cerebral asymmetry in prefrontal (Tallal et al., 1994) and posterior temporal cortical areas (Tallal et al., 1994; Leonard et al., 1993). Furthermore, functional imaging studies suggest that these anatomical differences are related to the inability of L-LD individuals to activate these areas (Neville et al., 1993; Rumsey et al., 1992; Hagman et al., 1992). Interestingly, in each of these studies, both the anatomical and functional measures show that these group differences were directly related to performance on tasks requiring the analysis of rapid acoustic change, not whether the information was verbal or non-verbal, suggesting that impaired temporal processing abilities may underlie the neural-based processing deficits of L-LD children.

Importantly, these data suggest that it may be the temporal acoustic characteristics of these phonemes that are initially coded in the brain, and that, with experience, these salient temporal codes come to be represented as categorical units. These studies provide novel insights into how
experience and learning directly affect neural processing, which may in turn lead to the storage of the representation of repetitive acoustic stimulation, such as speech sounds.

Learning and processing of speech and language

Early in life, infants discern differences between all the phonetic units used in the world’s languages, and demonstrate exquisite sensitivity to acoustic change in the region of the boundaries between phonetic categories (Eimas et al., 1987). Because non-human animals perceive the same discontinuities in speech, infants’ initial abilities have been attributed to more basic auditory processing mechanisms rather than ones that evolved specifically for language (Kuhl, 1991). By 12 months of age, infants fail to discriminate the foreign-language contrasts they once distinguished (Werker and Polka, 1993). Adults’ abilities are greatly reduced – they often find it difficult to perceive differences between sounds not used to distinguish words in their native language (Werker and Polka, 1993). Adult native speakers of Japanese, for example, have great difficulty discriminating American English /r/ and /l/ (Best, 1993), although Japanese infants do make this distinction (Tsushima et al., 1994).

Speech production follows a similar pattern. Regardless of culture, all infants progress through a set of universal stages during the first year (Ferguson et al., 1992). By the end of the first year, however, the utterances of infants reared in different countries begin to diverge, reflecting the ambient language (De Boysson-Bardies, 1993; Vihman and De Boysson-Bardies, 1994). In adulthood, the speech motor patterns that contribute to one’s ‘accent’ are very difficult to alter (Flege, 1993).

Work by Kuhl (1994) has produced an effect which helps to explain how language experience affects speech perception and production. The effect shows that linguistic experience alters the perceived distances between speech stimuli; in effect, experience ‘warps’ the perceptual space underlying speech. The end result is that perceptual categories are formed, ones that mirror the phonological categories of the ambient language. The phenomenon – called the ‘perceptual magnet effect’ – is demonstrated in experiments using phonetic ‘prototypes’, the best or most representative instances of phonetic categories. The experiments show that the best instances of phonetic categories function like ‘perceptual magnets’ for other sounds in the category. When listeners hear a phonetic prototype and attempt to discriminate it from sounds which surround it in acoustic space, the prototype displays an attractor effect on the surrounding sounds. It perceptually
pulls other members of the category towards it, making it difficult to hear differences between the prototype and surrounding stimuli (Kuhl, 1994).

These findings have been incorporated in a three-step theory of speech development, called the native language magnet (NLM) theory. NLM describes infants’ initial state as well as changes brought about by experience with language. It explains how infants’ developing native-language speech representations alter both speech perception and production. Phase 1 describes infants’ initial abilities. Infants’ abilities at this stage do not depend on specific language experience. The boundaries initially structure perception in a phonetically relevant way. However, they are not due to a ‘language module’ but to more basic auditory perceptual processing mechanisms. By six months (phase 2), infants have heard hundreds of thousands of instances of particular vowels (K. Gustafson, unpublished data). According to NLM, infants represent this information in memory in some form. Moreover, the distributional properties of vowels heard by infants raised in, for example, Sweden, America, and Japan differ. Linguistic experience has produced stored representations which mirror the vowel system of the ambient language. Language-specific magnet effects, produced by the stored representations, are now exhibited by infants. Phase 3 shows how magnet effects recursively alter the initial state of speech perception. Magnet effects functionally erase certain boundaries – those relevant to foreign but not native languages. At this stage, a perceptual space once characterised by simple boundaries has been replaced by a dynamically warped space dominated by magnets.

The important point is that infants at six months of age have no awareness of phonemes or the fact that sound units are used contrastively in language to name things. Yet the infant’s perceptual system has organised itself to reflect language-specific phonetic categories. At the next stage in linguistic development, when infants acquire word meanings by relating sounds to objects and events in the world, the language-specific mapping that has already occurred in their perceptual systems will greatly assist this process.

Infants learn to produce the sound patterns of language by listening to native speakers and imitating the sounds they hear. Speech motor control is a complex process, but by adulthood, people possess detailed information about the consequences of speech movements on sound. When do infants forge the perceptual-motor link? Recent studies suggest that the link is in place much earlier than 30 months if it is tapped by an appropriately sensitive measure. Infants’ utterances were recorded at 12, 16, and 20 weeks of age while the infants watched and listened to a video recording of a woman producing a vowel, either /a/, /i/, or /u/. Infants watched the video for 5 min on each of
three consecutive days, after which their utterances were analysed both perceptually (phonetic transcription) and instrumentally (computerised spectrographic analysis). It is found that the areas of vowel space become progressively more tightly clustered at each age.

Infants also imitated the vowels they heard. The total amount of exposure was only 15min yet this was sufficient to alter their vowel sounds. If 15 min of laboratory exposure to a vowel is sufficient to influence infants’ vocalisations, then listening to ambient language for weeks would be expected to provide a powerful influence on infants’ speech production. This data suggest that infants’ stored representations of speech alter not only infant perception, but production as well, serving as targets that guide motor production. Stored representations are thus viewed as the common cause for both the tighter clustering observed in infant vowel production and infant vowel perception (Kuhl and Meltzoff, 1995a, 1995b).

Mehler et al. (1988) demonstrated that French four-day-old newborns could discriminate French and Russian utterances. This ability remained when these utterances were low-pass filtered (erasing information above 400 Hz) to remove all phonetic cues and only the prosodic structure (namely global properties of the utterance like rhythm and melody) remained. Moreover, these newborns could discriminate between English and Spanish, two unfamiliar languages. By two months of age, a clear behavioural evolution was noticed – whereas two-month-old infants could discriminate a novel language from their maternal language, they no longer could discriminate two novel languages from each other. This suggests that two-month-old infants have established a template for their own language against which they evaluate all linguistic input. Apparently, speakers of different languages not only master different inventories of phonemes, but also segment the speech stream in a language-specific way. Mehler et al. (1981) and Segui (1984) proposed that the syllable plays a central role in the segmentation of speech, and that it is specific to the language.

Learning commences prenatally with the more prosodic aspects of language. Studies suggest that by the time infants are born, exposure to sound in utero has resulted in a preference for native-language over foreign-language utterances. Research has also demonstrated that the mother’s voice and simple stories she read during the last trimester are recognised by infants at birth. Estimates indicate that a typical listening day for a two year old includes 20 000-40 000 words. Speech addressed to infants is unique. However, the motherese pattern of speech, with its higher pitch and expanded intonation contours, is probably not necessary for learning. However, the context in which language is presented to the child, both its auditory and visual characteristics (greatly exaggerated facial expressions), fix infant attention on the talking caretaker (Burnham,
1993). The fact that linguistic input is accompanied by acoustic features that not only attract infant attention but mark its significant features probably helps infants learn.

When information about speech is retained, what form does the representation take? Two possibilities have been discussed in the adult literature on cognitive categories (Estes, 1993). The first assumes that people form some abstract version (a ‘prototype’) that characterises the category as a whole. The second, ‘exemplar-based’ model of categorisation, assumes that individual exemplars are stored and retrieved. Research by Miller et al. (1994) on adults shows that the location of best exemplars of a phonetic category shifts with changes in context such as the rate of speech, suggesting that speech representations are context-specific. Moreover, data collected by Pisoni et al. (1993) on the effects of talker variability, suggest that adult listeners encode fine details about the voice of the talker who produced the utterance, and that listeners’ subsequent recognition of speech information spoken by the same talker is improved. These data suggest either exemplar-based representations that contain context- and talker-specific instances or a number of prototype-based representations that are themselves context- or talker-specific.

Speech perception has classically been considered an auditory process. This belief has been modified by data showing that speech perception is strongly affected by the sight of a talker producing speech. One of the most compelling examples of the polymodal nature of speech is auditory-visual illusions that result when discrepant information is sent to two separate modalities. One such illusion can be demonstrated when auditory information for /b/ is combined with visual information for /g/. Perceivers report the phenomenal impression of an intermediate articulation (/da/, /tha/, or /za/) despite the fact that this information was not delivered to either sense modality. Recent data suggest the robustness of the effect by demonstrating it is maintained even when the cross-modal information cannot have derived from the same biological source, as when a male face is combined with a female voice (Green et al., 1991).

Moreover, when native speakers watch and listen to incongruent audio-visual speech signals pronounced by a foreign speaker, they show increased auditory-visual effects – greater numbers of illusory responses occur (Kuhl et al., 1994; Sekiyama and Tohkura, 1993). Kuhl et al. (1994) interpret these data as reflecting the fact that the auditory information in foreign speech does not match the stored representations of native-language speech – when this occurs, visual information may be more informative. The data support the idea that speech representations are polymodally mapped. Furthermore, young infants demonstrate knowledge about both the auditory and visual
information contained in speech, supporting the notion that their stored speech representations contain information of both kinds.

Various imaging techniques (PET, MRI, fMRI, and MEG) have not yet been applied to phonetic processing in infants, although adult studies are beginning to appear. However, high-density event-related potentials (ERPs) have recently been used to study word processing in young infants. In this study of two-month-olds, Dehaene-Lambertz and Dehaene (1994) presented infants with strings of syllables (/ba/ or /ga/), which were either identical or contained one deviant syllable. They observed two distinct peaks in electrical activity: peak 1, which occurred within 290 msec of the onset of the syllable, was insensitive to phonetic changes; peak 2, which reached its maximum about 390 msec after syllable onset, showed significant change when the deviant syllable was presented. Thus, a single instance of a deviant syllable was recognised in less than 400 msec in the infant brain. Future work will surely be aimed at mapping the brain changes which accompany language learning using these techniques of modern neuroscience.

B.2.5 Plasticity and organisation of action

Plasticity of sensorimotor and other cortical representations

The basis of the anatomical and functional organisation of sensory systems is the orderly representation or map of the peripheral receptor surface in an ascending series of connections within the central nervous system (CNS). The mammalian somatosensory system has two major ascending components, the trigeminal system and the dorsal column system, which are anatomically separate in the brainstem but are contiguous in ventroposterior thalamus and the primary somatosensory cortex (S1). The face is the receptor surface of the trigeminal system, whereas the remainder of the body is monitored by the dorsal column system (O’Leary et al., 1994). Apart from the mentioned map for the sense of touch, there is a similarly orderly representation of the muscles on the surface of the motor cortex, as shown in Figure B.6. It has also been shown that the retina is mapped not once but over and over in the cortex – the posterior third of the cerebral cortex is concerned with the multiple mapped representations of visual space (Gregory, 1987).
A question of fundamental importance is why in nature a great deal of trouble has gone into designing a set of genetic instructions which ensured that the retina of the eye and the surface of the body are represented on the surface of the brain in an orderly map and not higgledy-piggledy? The most plausible explanation concerns a well-known physiological phenomenon called lateral inhibition. For example, in the eye itself, adjacent differences in the brightness of the image are given prominence in the nerve signals leaving the eye. This is accomplished by a system of lateral inhibitory connections in the retina which ensure that nerve-cells tend to inhibit their immediate neighbours. In an area of uniform illumination, all cells are equally excited by the light and equally inhibited by their neighbours. However, where there is a sharp difference in illumination, as at the image of a contour, the highly illuminated cells exert a powerful inhibition on their neighbours in the shade, and the difference in signals sent by the two groups of cells is enhanced. Lateral inhibition cannot create something out of nothing, but it can enhance one feature of the visual image at the expense of another. Lateral inhibition of the kind just described ensures that edges and contours are prominently coded in the signals from the eye.
There is now excellent evidence from physiology and anatomy that lateral inhibition works in the brain as well as in the eye, and this provides the major reason for the existence of a map of the retina in the cortex of the brain. If the differences in illumination of adjacent parts of the eye are to be given further note in the cortex, then the connections between neurons concerned with the two adjacent parts of the image should be close together. In a map, they are as close together as possible, and lateral interactions will be maximally efficient. If there were no map at all, so that neurons concerned with adjacent parts of the image were often far apart in the brain, the problem of interconnecting the neurons would be formidable and the average length of a connection could be much greater. In the map of the sensory surface the lateral interconnections between cells can be local, and anatomy has shown this to be so.

However, why are there many maps rather than just one? The answer is really the same. Inhibitory connections between neighbouring neurons of the cortex are believed to be involved in coding many attributes of the visual image, such as colour, movement, disparity, orientation, size, and spatial periodicity. If all of this were to be attempted within one map, the local interconnections would again have to be longer and the problem of interconnecting the right cells would increase. By having many maps, each of which is small and contains neurons concerned only with one or a few of the stimulus attributes just mentioned, the lateral interconnections can be kept as short as possible and the problem of interconnecting the right type of cell is minimised. Although the different sensory qualities of the visual scene may initially be coded in separate visual areas, visual perception is unitary not fragmented, which means that the timing of the activity of cells in different visual areas must be precisely co-ordinated. If one looks at a moving, spinning coloured object and the neurons signals in one visual area were to be out of phase with all the others, some distortion should occur in what is seen. Indeed, fever, toxicosis, and brain damage may all lead to temporary visual perceptual dislocations (Gregory, 1987).

There is considerable evidence indicating that the functional organisation of S1 in adults (Figure B.6) can be altered by a variety of experimental approaches, most prominently by peripheral injury. In addition, several studies have added to the growing body of evidence that functional reorganisations may also occur on a moment-to-moment basis to help an animal adapt to a continuously changing environment. For example, the cortical representation of the reading finger in Braille readers can be shown to expand relative to their non-reading index finger, and the index fingers of sighted control subjects. Furthermore, researchers used magnetoencephalography (MEG), a non-invasive brain imaging technique, to map the somatotopic representation of the hand area in S1 of normal adult humans and in patients before and after surgical separation of webbed
fingers (syndactyly) (Mogilner et al., 1993), and found that the cortical representation of the hand and the individual fingers expands and its topography is reorganised within a few weeks after surgery.

A substantial body of information has also demonstrated that neuronal properties in auditory, visual and motor circuits can be reorganised in the adult mammalian CNS, either in response to lesions of peripheral or central structures or by experience alone (Donoghue, 1995). The majority of studies concerning sensory and motor plasticity have targeted the cerebral cortex, perhaps because of researchers’ fascination with this structure or possibly because of its experimental accessibility. However, as has been repeatedly pointed out in the literature, reorganisation of cortical cell properties or maps may reflect changes which occur at lower levels, which are then imposed upon the cortex. Reorganisation of motor cortex maps could occur in the subcortical structures through which the information passes before reaching its final motor output. For the somatosensory system, it is clear that nerve damage can produce rather complete reorganisation within the thalamus (Garraghty and Kaas, 1991). These sets of findings continue to raise the issue to what extent the plasticity reported in cortical representations actually occurs within the cortex itself.

A factor contributing to reorganisation is the ability of intracortical horizontal connections to grow (axonal growth) into the deafferented area (Darian and Gilbert, 1994). The limits to sprouting or its ubiquity remain to be determined – compared with the developmental period, the potential for growth in most adults is probably substantially more limited. A particularly intriguing finding is that introduction of nerve growth factor can re-introduce ocular dominance plasticity in adult cortex (Gu et al., 1994), suggesting that this process, which is usually limited to developmental stages, remains available in the adult. Two substrates for spreading information are known: sprouting of new connections and unmasking of existing ones.

Despite the growing body of evidence for some morphological restructuring in the adult brain, the regulation of the efficacy of existing synapses appears to be a major route for regulating the form of sensory and motor representations – two general candidate mechanisms exist: activity-dependent modification of individual synapses and regulation of postsynaptic cell excitability. Of the candidate mechanisms for regulating central organisation, activity-dependent synaptic modification, generally dubbed long-term potentiation (LTP) or long-term depression (LTD), seems to be widely available to sensory and motor pathways (Bear and Malenka, 1994). Although LTP/LTD is an attractive mechanism for regulating synaptic efficacy, changes in postsynaptic cell excitability can also participate in the dynamics of neural representations. In the cerebral cortex, alterations in
membrane excitability (by changing K+ conductances) accompany conditioning-induced changes in motor responses. Even though such a generalized postsynaptic change cannot preserve the synaptic specificity found with LTP/LTD, it can nevertheless alter the effectiveness of transmission among cells. These global changes in excitability may be ideal to ‘set up’ neurons for more synaptic-specific modification by lowering synaptic modification thresholds (Donoghue, 1995).

A final important development in understanding both the mechanisms and site of representational plasticity are studies of the molecular effects of experience or peripheral lesions upon neurons in the reorganising areas. These manipulations can have striking effects upon transmitter levels (Carder and Hendry, 1994), receptors (Huntsman et al., 1995; Hendry et al., 1994) and gene expression (Huntsman et al., 1995; Chaudhuri et al., 1995; Baekelandt et al., 1994). These approaches hold considerable promise in understanding the sequence of events that unfold during neural reorganisation and may provide insights into how to manipulate these systems to promote recovery from disease and injury or to overcome learning or memory dysfunction (Donoghue and Sanes, 1996).

Functions and structures of the motor cortices in humans

Humans and non-human primates have many motor cortices; however, it is not yet clear exactly how many there are. Two major issues that remain unresolved are the exact criteria for defining a motor area and how to apply these criteria to candidate motor areas. Most researchers in the field will agree on certain criteria, such as that a motor area should both have projections to neurons in the anterior horn of the spinal cord or in the motor nuclei of the brain stem and have full representation of the somatomotor apparatus (voluntary motor control of the whole body from face down to toes). At present, however, it is practically impossible to verify projections to the spinal cord in humans. Even though it is possible to assess, in gross, whether an area contains a full representation of the somatomotor apparatus, the lack of a somatotopical organisation makes interpretation of data difficult. Another generally accepted criterion is that a motor area should be active whenever voluntary movements are planned or executed, but it should be only active occasionally under other circumstances. Even on the basis of the latter two criteria, which are now possible to apply to human studies (through functional imaging techniques), there is hardly any single motor subdivision in which a full representation of the somatomotor apparatus has been demonstrated and for which the exclusive motor function have been demonstrated, with the exception of the primary motor cortex (M1) and possibly of the supplementary motor area (SMA) (Roland and Zilles, 1996).
One of the major problems facing work on human motor areas is that the microstructure (cytoarchitecture, myeloarchitecture and receptor architecture) of many areas of the brain is not yet known. Authors often optimistically refer to localisations of fields of activation within Brodmann's cytoarchitectural areas on the basis of gross morphological resemblances between the two-dimensional cartoon of Brodmann (1909) and their three-dimensional images, forgetting that Brodmann's map did not consider the large variations in extent and topology that exist among human brains (Filimonoff, 1932; Roland and Zilles, 1994; Rajkowska and Goldman-Rakic, 1995) and that the human cerebral cortex is virtually devoid of true macroscopical landmarks delimitating architecturally defined areas (Roland and Zilles, 1994). What are pertinent in these cases are three-dimensional maps of the likelihood that a functional field is located within a certain microstructurally defined motor area.

Within M1, the centres of gravity (or, alternatively, the points of maximal activation) for various functional activations appear to be ordered somatotopically (Kawashima et al., 1996b; Matelli et al., 1993; Geyer, et al., 1996; Grafton et al., 1993; Fried et al., 1994). However, the activations associated with movement of one body part overlap with those of adjacent body parts (Sanes et al., 1995). For example, thumb, index, ring-finger and wrist representations overlap each other by 40-70% (when measuring the volume of activation). This picture is further complicated by new observations that the number, localisation and extent of activation in M1 associated with movement of one body part are modifiable by learning (Kawashima et al., 1994; Pascual-Lenon et al., 1994; Karni et al., 1995).

According to Karni et al. (1995), the volume of the field associated with the performance of a learned motor sequence is larger than that associated with a new sequence. Whatever the future view of plasticity, shifts and multiple representations of single body parts may be, this study is incompatible with the view of a single highly ordered map of the executive sites of the somatomotor apparatus. Karni et al. (1995) suggest that the rough scheme of fields activation associated with the face, hands, arms and legs prevails, but that the overlap of representations of more localised parts of the somatomotor apparatus is considerable (Penfield and Boldrey, 1937), and that the maps themselves are more dynamic than previously assumed.

Complex sequences of movements do not seem to activate different motor areas from those activated by simple flexion-extension of one finger. The premotor cortex (PM), SMA, and M1, and occasionally the cingulated motor area (CMA), are activated by such complex motor sequences.
Distributed motor processing in cerebral cortex

While the desire to move can originate in the absence of overt environmental stimuli, it is much easier to study movement generation when it is linked to known cues. The processing by successive CNS structures of the sensory cues leads to the action embodied in the desired movement. The trajectory of the movement is a function of the desired behaviour. Inherent in the description of the movement is the idea of parameters – physical quantities which can be measured and represented in the activity of neurons. Typically, sensory parameters are those aspects of the environment which act on the individual, whereas motor parameters are measurements of the actions taken by the individual to act on the environment. As soon as movement occurs, both types of parameters change together and the nervous system processes them simultaneously. Movement changes the environment, altering the sensation used to generate the next portion of the movement. This cyclic interaction blurs the boundary between sensory and motor, a distinction which becomes more artificial when considering neuronal connections which are a few synapses away from the periphery (Schwartz, 1994).

The widespread observations of broad tuning in different neuronal systems and the ability to detect external parameters using population algorithms (for example, combining the weighted contributions of a number of neurons) in these systems suggest the exciting possibility that researchers may soon begin to understand how information about the external world is transformed to a state where it is used to trigger and control movement. The flow of visual information to motor cortical areas is shown in Figure B.7 (Schwartz, 1994).
Figure B.7 Summary of connection between visual cortical areas (V1, V2) and primary motor cortex (M1). (Cb = cerebellum; Cing = cingulated cortex; PN = pontine nuclei; pSMA = pre-SMA; PM = premotor cortex; VPLo = oral portion of ventral posterior lateral thalamus) (Schwartz, 1994)

Visual information has been described as projecting through the neocortex in separate pathways. Spatial information flows from layers 4b and 6 of cortical areas V1 and V2 to the medial temporal area (MT) in a dorsal pathway. The pathway then projects to the medial superior temporal area (MST). The ventral pathway is believed to involve cortical areas V1, V2, V4, and the inferotemporal area (IT), structures which are involved in object identification. It is interesting to note that recent anatomical studies suggest that there may be a third major visual pathway, which projects to the parietal cortex via the V3-V3A complex, and is probably concerned with stereopsis. These results imply that the initial binocular view, before movement onset, provides important information about size, shape, and distance of an object for smooth and accurate execution ofprehension movements and that binocular vision is an important source of feedback information (Sakata and Taira, 1994).

Cells in MT are broadly tuned to the direction of visual motion. Dot patterns moving in different directions generate discharge patterns that when mapped to direction result in tuning functions which are very similar to those found in the motor cortex for different directions of arm movement (Snowden et al., 1992). Whether these neuronal responses can be summed to produce population responses or operate individually in a ‘winner-take-all’ mechanism (Salzman and Newsome, 1994)
is presently somewhat controversial. MT cells project to MST. Cells in the dorsal portion of this area, MSTd, respond to more complex motions composed of rotation, expansion and translation (Sakata et al., 1985). These are components of ‘visual flow’, that is, information experienced when moving through the environment. The responses of individual cells are related to combinations of the three components and have been fit with coarse tuning functions in a way that suggests that many cells respond simultaneously as this type of visual information is processed (Graziano et al., 1994). For instance, when the random dot patterns moved in a spiral motion (combination of expansion/contraction and rotation), many of the cells responded with high discharge rates and would respond to other motions with lower rates. Displaying these patterns in other portions of the receptive field did not have a great influence on the cellular responses, suggesting that they were position invariant. Positional invariance is not a property of MT cells and probably is a function of the integration of MT inputs in MST. In addition to visual flow, MST discharge could be used to process information during manipulation because these cells respond when an object in the foreground moves against a background (Schwartz, 1994).

MST and MT project to cortical area 7a and to areas within the intraparietal sulcus (VIP and LIP). Cells in these regions respond broadly to spots of light moving from the periphery towards the fovea in different directions. These cells probably play a role in the visual guidance of the hand as it approaches a foveated area. Many of the areas in and around the intraparietal sulcus project to the frontal cortex. Some of these projections terminate in the premotor areas. Area 7m receives input from parietal-occipital cortex (PO), an area which processes visual stimuli in the peripheral field. MIP cells have visual and motor responses. Cells in dorsal premotor cortex (PMd) are broadly tuned to the direction of arm movement with the same tuning function as cells in M1. As with M1 activity, a population algorithm applied to PMd activity formed accurate neural trajectories. One characteristic of these cells is that a subpopulation appeared to be inhibited for curved trajectories. Cells in ventral premotor area (PMv) receive their major cortical input from the fundus of the intraparietal sulcus – area VIP. This parietal area contains cells responsive to visual motion. PMv cells may be gaze specific, in that their activity changes when static stimuli are fixated in different parts of the visual surround, but seem to be gaze independent for moving stimuli. Area 7 output can also reach motor cortical areas through a cerebellar circuit. There is a large projection from this region to the pontine nuclei, lateral cerebellar cortex, dentate nucleus, ventrolateral thalamus, and the primary and premotor cortices. Cerebellar cells are sensitive to moving visual stimuli and are also broadly tuned to the direction of arm movements (Schwartz, 1994).
In conclusion, there is no direct route from visual cortical areas to M1. There are, however, several pathways through ipsilateral cortical sites and subcortical structures from visual areas around the intraparietal sulcus to M1. This route is characterised by motion processing of visual stimuli. Visual information associated with object recognition may reach the motor cortex more indirectly from a pathway projecting from IT to prefrontal, then premotor and supplementary cortices. However, caution must be used when interpreting this type of anatomical routing. First, almost all of the premotor cortical areas have direct spinal projections independent of M1. Second, functional imaging studies repeatedly find that these pathways do not function independently. Third, it is difficult to assign discrete functions to particular anatomical substrates, as the properties of their cell activity during movement are similar. With these problems, how might one gain a realistic understanding of the processing responsible for the development of volitional action based on visual stimuli? It is likely that parameters used in this process are represented with broad tuning functions in every structure of the pathway. One approach would be to record the activity of many cells simultaneously from different structures in the pathway. Each cell could be mapped analytically to an n-dimensional space based on its response to a variety of parameters. The dimensions of this space can be any aspect of the experiment which can be measured. If this mapping were done with a large sample from many of these structures, a time-series of n-dimensional population vectors through this space, calculated at intervals throughout the task, should reflect accurately the processing subserving the volitional action. If cells within a structure are involved in movement generation in a similar way, they should form clusters in this space and the evolution of these clusters throughout the task could explain how cellular activity in one structure interacts with that of another (Schwartz, 1994).

Role of the cerebellum and basal ganglia in movement control and representation of temporal information

To understand how the brain controls movement, it makes sense to study the cerebellum as it sends large projections directly, or through relays, to every major motor structure in the brain. The cerebellar influence on these motor structures reflects a considerable investment of brain resources as the cerebellum contains about $5 \times 10^{10}$ neurons, which is in the same range as the number of neurons estimated to be in the whole cerebral cortex, $2.2 \times 10^{10}$. Understanding the cerebellum seems within reach because its internal structure is simple and surprisingly consistent throughout. In spite of the apparent importance and tractability of the cerebellum, researchers still have only basic information about what it does and how it does it. For example, the cerebellum is not indispensable for movement because cerebellar-injured patients can still move the parts of their
bodies which are influenced by the damaged part of the cerebellum. However, the movements of these parts are slow, inaccurate and jerky. It appears, therefore, that the cerebellum is necessary for fast, accurate, and smooth movements. The cerebellum is also necessary for movement adaptation, which is the unconscious adjustment that the brain makes to correct the size or direction of a movement so as to make the movement accurate after something has made it inaccurate. Movement inaccuracies occur naturally as a result of growth, aging, trauma, or disease (Robinson, 1995).

For example, a baby learns to sit up when the organs of balance in the internal ear have established reflexes with the motoneurons working the muscles of the trunk and limbs. The basic reflexes for standing, crawling, sitting, walking, and running depend on built-in circuits of neurons in the spinal cord, connected to circuits in the centre of the cerebral hemispheres and the cerebellum. The cerebellum is kept informed about the activity of all muscles, about acceleration and deceleration affecting the whole body, and about all other inputs to the CNS. As soon as the spinal cord starts performing a movement, impulses are sent up to the cerebellum and continue to be sent throughout the movement, so that the cerebellum is kept informed on how the movement is progressing. Therefore, the cerebellum is an error-measuring device which compares the actual performance with the programme which it receives from the cerebral cortex. Most movements are programmed by the brain and they use lower-level reflex arrangements as their components. When a person performs touch-typing, the impulses leave the brain, programmed in time and place to hit the keys in order. Once the impulses are on their way, they cannot be interrupted. The person knows when he/she is going to type the wrong letter or the letters in the wrong order, and out they come, wrong; the knowledge comes too late to interrupt the planned movements (Gregory, 1987).

The fourth dimension of the world, time, has tended to be the forgotten stepchild in many theories of perception and motor control. Even though actions evolve over time, it is not mandatory to postulate that temporal information is represented or regulated in an explicit manner. Variation in the speed and duration of a reaching movement might be an emergent property of the rate at which muscle units are recruited. Therefore, temporal regularities in sequential actions may not reflect direct control processes, but may arise because of the complex dynamics of the neuromuscular system. Nonetheless, many phenomena suggest the existence of an internal timing system in which temporal information is explicitly represented. It has been hypothesised that the cerebellum operates as a specialised module for timing (Ivry, 1993; Ivry, 1997). The basal ganglia have also been suggested to be a key component in an internal timing system.
Endogenous rhythms are observed at multiple levels of the nervous system, operating over many different time scales. In motor control studies, researchers have described how central pattern generators, composed of small networks of spinal neurons, produce complex patterns of locomotion. The existence of such periodic mechanisms motivated models of time perception and production centred on a putative central pacemaker. In these models, the temporal pacemaker is composed of two parts: an oscillator and a calibration unit. The oscillator produces an output at a constant frequency. The calibration unit re-scales this base frequency as a function of external influences and task demands. Together, this pacemaker provides flexible timing information.

Using a series of interference tests in time production and perception experiments, estimates of the oscillatory frequency have converged on a value of approximately 49 Hz (Ivry, 1996). This value is of interest given recent neurophysiological evidence suggesting that oscillatory brain activity near 40 Hz might serve as a mechanism for integrating activity across different neural regions (Gray and Singer, 1989; Joliot et al., 1994; Poeppel, 1996). By allowing for flexibility in the calibration unit, pacemaker models can account for why the subunits of an action retain their proportional timing when that action is performed at different overall rates (Viviani and Laissard, 1996; Collier and Wright, 1995).

Recent challenges to this idea come from network models in which time is distributed across a set of neural elements, with the different elements not only providing multiple timing mechanisms, but also that these mechanisms are linked to particular task domains. For example, according to this view, there would be a set of timing elements to regulate tapping at different rates with one limb, with this organisation repeated for other limbs. An analogy can be drawn here to the way cells in the visual cortex are tuned to edges at different orientations, although physiologists have not observed chronotopic maps in any brain area to date. However, this distributed representation could be restricted to a single neural structure such as the cerebellum. In this view, the representation of time might be one defining property of the computational capability of that structure, although the exact elements recruited would be task-dependent. This hypothesis seems more biologically plausible in comparison to models postulating a single internal clock (Ivry, 1996).

A major shift is under way in work on the basal ganglia (Graybiel, 1995). After years of painstaking work to build up knowledge of the basic anatomy, neurochemistry and cell physiology of the system, studies are now increasingly being directed towards functional issues and towards testing hypotheses about basal ganglia function. At the same time, there is renewed emphasis on trying to understand the functions of the basal ganglia in terms of their interactions with the neocortex and other structures as parts of distributed neural circuits. This activity is attracting
many researchers to work on the basal ganglia, including computational scientists. There is no consensus on ‘the’ function of the basal ganglia, but work with techniques ranging from imaging to recording and chemical inactivation of basal ganglia sites in primates and rodents points to the basal ganglia as part of a highly dynamic neural system involved in adaptive control of action, not only in the motor sphere, but also at the level of planning and cognition. The key functional theme is that the basal ganglia are involved in one or more stages of the building up, storage, decoding, retrieval, and expression of behavioural action plans through collaboration with the neocortex, thalamus, and the limbic system. Evidence suggests extreme specificity in the neural connections interrelating the basal ganglia, cerebral cortex and thalamus. Adaptive control of behaviour may centrally depend on these circuits and the forebrain and midbrain evaluator-reinforcement circuits which modulate them.

The organisation of behaviour sequences, whether learned or innate, may critically depend on neural processing in the striatum (part of the basal ganglia), which not only receives massive cortical inputs relevant to sequence generation, but is also in a position to act on the cortex and on pattern generators in the brainstem. But what is the nature of sequence-related neural processing in the striatum, and how does it differ from and functionally integrate with sequence-related processing in the SMA and other cortical areas? Recent models do raise timing as a critical feature of neural processing in the striatum (Graybiel, 1995).

First, evidence suggests that many different cortical inputs need to converge on individual striatal projection neurons to bring these spiny neurons to firing threshold (Wilson, 1995), and, further, that cortical inputs to these neurons are re-mapped systematically into convergent or divergent modules (Graybiel and Kimura, 1995; Parthasarathy et al., 1992; Flaherty and Graybiel, 1994; Cowan and Wilson, 1994). This sets up a situation in which the striatal projection neurons could act as modifiable coherence-detection elements (Graybiel, 1994; Parthasarathy et al., 1992; Flaherty and Graybiel, 1994; Cowan and Wilson, 1994): the hard-wired patterns of convergence of their cortical (or other) inputs may thus represent a template for neural processing, but activity-dependent temporal coordination of the inputs may determine their spike activity and hence striatal output activity.

Second, there are several sets of striatal interneuron that could act differentially to modulate the spatiotemporal dynamics of striatal input-output processing. One set, the cholinergic interneurons, is thought to correspond to the tonically active neurons (TANs), whose firing can be modified during behavioural learning. It has been suggested that these interneurons could serve a 'motor-
Mental motor imagery

Most of people’s actions are driven indirectly by internally represented goals, rather than directly by the external environment (Jeannerod and Decety, 1995). Until recently, the existence and structure of such motor representations were inferred from the duration and timing of a reaction, or from the pattern of executed movements. Now, however, a more direct approach has been adopted which exploits the unique ability of human subjects to image and simulate actions consciously. Motor imagery is a cognitive state which can be experienced by virtually everyone with minimal training. It corresponds to many situations experienced in everyday life, such as watching somebody’s action with the desire to imitate it, anticipating the effects of an action, preparing or intending to move, refraining from moving, or remembering an action. Using motor imagery as a means of analysing covert processes seems justified by previous work on mental imagery in other modalities. For example, visual imagery engages many of the mechanisms and neural structures employed in visual perception (Farah et al., 1990; Kosslyn et al., 1993; Roland and Gulyas, 1994). It seems logical, therefore, to look at the motor system for the same direct continuity between mechanisms for the representational stages of an action and (action) performance.

Mental simulation of movement activates motor pathways. During motor imagery, muscular activity increases with respect to rest. When this is the case, electromyographic (EMG) activity is limited to those muscles which participate in the simulated action, and tends to be proportional to the amount of imagined effort. The fact that muscular activity is only partially blocked during simulation of movement suggests that motoneurons are close to threshold. However, EMG activity can also be quiescent. This does not necessarily contradict the link between motor imagery and muscular activity, as it merely reflect better inhibition of movement execution under certain conditions or in certain subjects (Jeannerod and Decety, 1995). It has also been shown that the size of the area responding to finger movements increases as simulated movements are repeated over training periods, in the same way as when actual movements are repeated. In addition, ‘concentrating’ on one hand muscle without activating it increases the effect of subthreshold magnetic stimulation of the cortical area corresponding to that specific muscle. Thus, there is a
selective enhancement of responsiveness to stimulation of motor cortical areas during motor imagery. In a study to determine to what extent people can distinguish their self-generated actions from those performed by others, it was found that subjects can recognise their own movements reliably better than change from their detailed kinematic properties. This result is evidence for a contribution of motor imagery and planning mechanisms to even perception and thus for a common coding in the brain for perception and action (Wexler, 2001).

These results raise the problem of the mechanism and the locus of motor inhibition during motor imagery. During motor preparation, the movement is blocked by a massive inhibition acting at the spinal level to protect motoneurons against a premature triggering of action – hence the decrease of spinal reflexes during the preparatory period and their re-increase shortly before the movement starts (Bonnet and Requin, 1982). During mental simulation, it is likely that the excitatory motor output generated for executing the action is counterbalanced by another, parallel, inhibitory output. The competition between two opposite outputs would account for the partial block of the motoneurons, as shown by residual EMG recordings and increased reflex excitability. It is not yet possible to identify whether this inhibitory output originates in the cortex or elsewhere.

The autonomic system (regulating heart rate, respiration rate and CO₂ pressure), normally not submitted to voluntary control, is also activated during motor imagery. Research results confirm that a large fraction of the fast increase in heart and respiration rates at the onset of exercise (both real and mental) is due to central factors rather than metabolic changes. Vegetative activation during preparation for effort is thus part of motor programming. It is timed to begin when motor activity starts, which represents an optimal mechanism for anticipating the forthcoming metabolic changes and shortening the intrinsic delay needed for heart and respiration to adapt to effort (Jeannerod and Decety, 1995). The possibility that these autonomic changes are a consequence of muscular activity can be ruled out, as there is no change in muscular metabolism during mental simulation.

Pioneering studies using two-dimensional regional cerebral blood flow (2-D rCBF) mapping or single photon emission computed tomography (SPECT) have emphasised the activity of several brain areas during motor imagery. Prefrontal areas, supplementary motor are (SMA), cerebellum, and basal ganglia are the main activated areas. Recent positron emission tomography (PET) studies reveal that brain activity is in fact influenced by the nature of the imagery task. Decety et al. (1994) instructed subjects to imagine themselves grasping visually presented three-dimensional objects with their right hand. Brodmann area 6 (pre-motor area) in the inferior part of the frontal
gyrus on both sides as well as area 40 in the contralateral inferior parietal lobule were strongly activated. Subcortically, the caudate nucleus (part of basal ganglia) was found to be activated on both sides and the cerebellum on the left side. Another focus of activity was observed in left prefrontal areas, extending to the dorsolateral frontal cortex. Finally, the anterior cingulated cortex was bilaterally activated.

In other studies, where the task consisted of repetitive, internally generated eye and hand movements, an additional activation of SMA was observed. Interestingly, comparison of externally and internally generated movements in the same subjects showed that SMA activation during simulated movements was more rostral than commonly observed during executed movements (Stephan et al., 1995). This finding reinforces the notion that SMA is divided into areas of different hierarchical status with different functional implications: the posterior zone is purely executive (the SMA proper), whereas the more anterior zone is more related to representational stages of action and to motor imagery.

Taken together, the results on the neural correlates of motor imagery provide a good basis for explaining the effects of ‘mental practice’, now commonly used by sportsmen for mentally rehearsing motor performance. Motor imagery and related states, such as observation of actions performed by others, produce a selective enhancement of neural activity in those motor pathways concerned with the simulated action. This leads to an increase in muscle strength and a decrease in the variability of movements. These results have important implications for the mechanisms of motor learning. As selective improvements in motor performance can be obtained in the absence of an increase in muscular activity (and therefore without re-afferent input from the muscle), they suggest that learning could be due to a purely central shaping of motor output (Jeannerod and Decety, 1995).
B.3 Social Cognition

B.3.1 Overview

The emotional and social development of humans is extraordinarily complex, involving a multifactorial interplay between genes, parental behaviour and the influence of culture. A ‘Theory of Mind’ (TOM) is a specific cognitive ability to understand other persons as intentional agents, that is, to interpret their minds in terms of theoretical concepts and intentional states such as beliefs and desires. It has been commonplace in philosophy to see this ability as intrinsically dependent upon linguistic abilities. After all, language provides people a representational medium for meaning and intentionality – thanks to language a person is able to describe other people’s and his/her own actions in an intentional manner. According to this view, the intentionality of natural language, that is, the suitability for expressing meanings and thoughts, is the key for understanding the intentionality of a person’s theory of mind.

A major challenge to this view came from studies on primate cognition and comparative psychology. Premack and Woodruff (1978) argued that experimental evidence of chimpanzees’ understanding of human behaviour could be interpreted as detection of intentions. Although other primatologists have challenged their experimental data, there is growing evidence showing that non-human primates have some intentional understanding of their social world (Tomasello and Call, 1997). The presence of such capacity in non-human (and thus non-linguistic) species leads to the conclusion that it is possible to investigate TOM as a biological endowment independently of language.

Humans are exceedingly social beings, but the neural underpinnings of social cognition and behaviour are not well understood. Studies in humans and other primates have pointed to several structures which play a key role in guiding social behaviours: the amygdala, ventromedial frontal cortices, and right somatosensory-related cortex, among others. These structures appear to mediate between perceptual representations of socially relevant stimuli, such as the sight of conspecifics, and retrieval of knowledge (or elicitation of behaviours) which such stimuli can trigger. Current debates concern the extent to which social cognition draws upon processing specialised for social information, and the relative contributions made to social cognition by innate and acquired knowledge (Adolphs, 1999).
Social cognition refers to the processes that subserve behaviour in response to conspecifics (other individuals of the same species), and, in particular, to those higher cognitive processes subserving the extremely diverse and flexible social behaviours which are seen in primates. Its evolution arose out of a complex and dynamic interplay between two opposing factors: on the one hand, groups can provide better security from predators, better mate choice, and more reliable food; on the other hand, mates and food are available also to competitors from within the group. An evolutionary approach to social cognition therefore predicts mechanisms for cooperativity, altruism, and other aspects of pro-social behaviour, as well as mechanisms for coercion, deception and manipulation of conspecifics (Adolphs, 1999). The former are exemplified in the smallest groups, in the bond between mother and infant; the latter in the largest groups by the creation of complex dominance hierarchies.

B.3.2 Social cognition, modularity and innateness

It is known that focal brain damage can result in impaired processing that is limited to highly specific categories. For instance, patients have been reported who are specifically unable to recognise, or to name, tools, animals, people, or a variety of other selective categories. There is thus very strong evidence that categories are, in some sense, mapped in the brain. This finding is predicted from the assumption of a few, very simple, local rules which specify how brains represent stimuli (Kohonen and Hari, 1999). In essence, local rules for organising neural tissue as a function of activity suffice to generate topographic representations of abstract stimulus categories. The categories which are abstracted emerge naturally out of the covariances of people's interactions with certain classes of stimuli in the environment. Thus, people typically interact with members of the class of animals in a similar way; that is, the similarity is greater among animals than it is to how people typically interact with members of the class of tools, or members of the class of people. Similarity in sensorimotor interaction can thus translate into functional and anatomical similarity in the brain (Solomon et al., 1999).

The above view suggests a strong component of experience and learning in such self-organised topographic maps. A different explanation comes from the view that there are innately specified modules in the brain for processing specific categories of knowledge. The evidence for this latter view is strongest from domains such as language, and it is the view that has historically been associated with the notion of ‘modularity’ (Fodor, 1983). As with many dichotomies, it is likely that both views are right, in the proper context, and recent interpretations suggest a softer version of ‘modularity’ that does not require a rigid set of criteria (Coltheart, 1999). It may well be that there
are domain-specific modules for processing certain kinds of information which are ecologically highly relevant and that would benefit from a particular, idiosyncratic processing strategy which does not apply to other kinds of information. That is, one would expect the brain to provide problem-specific structures for processing information from those domains in which there is a premium on speed and survival. Within, and beyond, such a module there might also be topographic mapping of the same domain. It is likely that domain-specific processing draws upon innately specified modules, as well as upon self-organised maps which emerged as a consequence of experience with the world (Adolphs, 1999).

Some rather basic attributes of stimuli, such as self-directed motion, bilateral symmetry, and presence of eyes might be processed similarly by different primate species, by mechanisms which are highly innately specified. However, there also seems little doubt that the class of social stimuli needs to be explored during development in order to be able to make more fine-grained distinctions – a development process which is likely to include parental behaviour and pretend play as critical aspects. The most plausible scenario, then, would view social cognition as relying on a neural architecture in which there is interaction between components which are innately specified and others whose operation emerges through experience in the context of a specific culture (Adolphs, 1999).

There have been two major sets of studies that first argued for neural system critical to social cognition in humans: social impairments following damage to the frontal lobe, and, more recently, social impairments in subjects with autism (Adolphs, 1999). Subjects with bilateral damage to the ventromedial frontal lobes show a severely impaired ability to function in society, despite an entirely normal profile on standard neuropsychological measures, such as IQ, language, perception and memory. Recent theoretical explanations propose that the ventromedial frontal cortices play an important role in associating emotional experience with decision-making in complex situations, especially situations in the social domain.

A second line of evidence that has been used to argue for the functional modularity of social cognition comes from a developmental disorder, childhood autism. Interest in the social cognitive abilities of subjects with autism was fuelled by the argument that autism features a disproportionate impairment in one specific aspect of social cognition – the ability to attribute mental states, such as beliefs, to others (Leslie, 1987; Baron-Cohen, 1995). While there is debate on the basic hypothesis, and while the link between autism and brain systems is also not well understood, the
data point towards neural components which appear to have a high degree of domain-specific function (Adolphs, 1999).

**B.3.3 Social judgment of faces in the amygdala**

People glean considerable social information from faces, and there is evidence to suggest that faces are processed in a relatively domain-specific fashion by neocortical sectors of the temporal lobe. Recent data suggest that it is a particular property of how people interact with faces that leads to the specific neuroanatomical processing seen, namely, that people need to become experts at distinguishing many exemplars which are visually extremely similar and yet socially highly distinctive (Gauthier et al., 1999).

High-level visual cortices in the temporal lobe project to the amygdala (Amaral et al., 1992), which has also received historical and recent interest concerning its role in processing emotionally and socially salient information from faces. Studies which have examined patients with damage to the amygdala (Adolphs et al., 1994; Young et al., 1995; Adolphs et al., 1999) have provided evidence that the amygdala is critical to recognise emotions from facial expressions, specifically certain negative emotions, such as fear. While these threads of research have pointed to a disproportionately important role for the amygdala in processing stimuli related to danger and threat, there are also findings, primarily from studies in animals, which suggest a more general role for the amygdala in processing emotionally arousing stimuli which are either pleasant or aversive. One recent theoretical view suggests that the amygdala, in both humans and animals, might subserve a more general role in allocating processing resources to biologically salient stimuli which are ambiguous, and about which additional information needs to be acquired, regardless of the valence of those stimuli (Whalen, 1999).

Given the above findings, it might be expected that the amygdala also make important contributions to higher-level social cognition, especially to those aspects of it which rely on recognising social information from faces. One important cue, direction of eye gaze in a face, has been shown to be processed by the amygdala (Adolphs et al., 1994). Other studies have examined the role of the amygdala in more social judgments through investigating subjects’ ability to judge how trustworthy or how approachable other people looked, from perceiving their faces. Such an ability draws on aspects of social recognition, as well as on social decision making. The role of the amygdala in processing stimuli related to potential threat or danger thus appears to extend to the complex judgments on the basis of which people regulate their social behaviour. Clearly, the cues which are
normally used to make such judgments will be complex, and there will be multiple strategies available to utilise them (Adolphs, 1999).

Observation of patients with complete bilateral amygdala damage suggests a common aspect to their social behaviour: they tend to be unusually friendly towards others, consistent with the idea that they lack the normal mechanisms for detecting individuals who should be avoided. Similar changes in behaviour are seen in non-human primates with selective bilateral amygdala damage (Emery et al., 1998; Meunier et al., 1996). On the other hand, the human patients do not appear to be as severely impaired in their social behaviour as do monkeys with similar brain damage. It may be that humans with amygdala damage, unlike other animals, possess additional mechanisms for social reasoning and decision making, and are able also to draw substantially on declarative knowledge encoded in language, resulting in partial compensation for their impairment (Adolphs et al., 1995).

A final important consideration concerns the role of the amygdala beyond recognition and judgment, to encompass such processes as attention and memory. It is clear from studies in animals that the amygdala contributes importantly to these processes (Holland and Gallagher, 1999), and its role extends well beyond a function restricted to recognising potential threat or danger; but such a possible role in humans is just beginning to be explored. For instance, emotionally (Bradley et al., 1992) or socially (Mealey et al., 1996) salient stimuli are remembered better by normal individuals, an effect that correlates with activation of the amygdala in functional imaging studies (Cahill et al., 1996; Hamann et al., 1999) and one whose function is impaired in patients with amygdala lesions (Adolphs et al., 1997).

Taken together, all the above findings argue that the amygdala is one component of the neural systems by which stimuli trigger emotional reactions, broadly construed. Such emotional reactions would include autonomic, endocrine and somatomotor changes in the body, as well as neurophysiological and neuromodulatory changes in brain function. Such multi-dimensional emotional responses would serve to modulate and to bias cognition and behaviour in important ways, as a function of the emotional and social significance of the stimulus which is perceived. This role for the amygdala may be of special importance for relatively fast, automatic evaluation of biologically important stimuli, and will no doubt function in parallel with other systems.
B.3.4 Social reasoning and decision making in the ventromedial prefrontal cortex

More recently, it has become clear that the frontal lobes, specifically their ventromedial sectors, are critical in linking perceptual representations of stimuli with representations of their emotional and social significance (Damasio, 1994). This function bears some resemblance to that of the amygdala outlined above, but with two important differences. First, it is clear that the ventromedial frontal cortices play an equally important role in processing stimuli with either rewarding or aversive contingencies; whereas the role of the amygdala, at least in humans, is clearest for aversive contingencies. Second, reward-related representations in the ventromedial frontal cortex are less stimulus-driven than in the amygdala, and can be the substrate of more flexible computations, playing a general monitoring role in regard to both punishing and rewarding contingencies (Schoenbaum et al., 1998).

The impaired social behaviour in humans with ventromedial frontal lobe injury is notable for an inability to organise and plan future activity, a diminished capacity to respond to punishment, stereotyped and sometimes inappropriate social manners, and an apparent lack of concern for other individuals, all in the face of otherwise normal intellectual functioning (Damasio, 1994; Adolphs, 1999). Particularly striking are the patients’ often gross lack of concern for the well-being of others and remarkable lack of empathy.

The role of the human ventromedial prefrontal cortex in decision-making has been explored in a series of studies using a task in which subjects had to gamble in order to win money. The key ingredient which distinguishes this task from other tasks of probabilistic reasoning is that subjects discriminate choices by feeling; they develop hunches that certain choices are better than others, and these hunches can be measured both by asking subjects verbally, and by measuring autonomic correlates of emotional arousal, such as skin conductance response. Subjects with damage to the ventromedial frontal cortex fail this task (Bechara et al., 1994), and they fail it precisely because they are unable to represent choice bias in the form of an emotional hunch (Bechara et al., 1997).

These findings are consonant with prior reports that subjects with ventromedial frontal lobe damage do not trigger a normal emotional response to stimuli, including socially relevant stimuli (Damasio et al., 1990), and support a specific hypothesis that has been put forth to explain the data, namely the somatic marker hypothesis (Damasio, 1994, 1996). According to this hypothesis, the ventromedial frontal cortex participates in implementing a particular mechanism by which people acquire, represent and retrieve the values of their actions. This mechanism relies on generating
somatic states, or representations of somatic states, which correspond to the anticipated future outcome of decisions. The function of these somatic states is to steer the decision-making process towards those outcomes which are advantageous for the individual, based on the individual’s past experience with similar situations. Such a mechanism may be of special importance in the social domain, where the enormous complexity of the decision space precludes an exhaustive analysis.

The ventromedial frontal cortex appears to play a key role in a second domain of high relevance to social cognition, namely social reasoning. Human reasoning strategies have been intensively investigated using the Wason selection task, the most popular experimental design for probing deductive reasoning. Findings also support a role for the ventromedial frontal cortex in guiding reasoning and decision making by the elicitation of emotional states which serve to bias cognition. While the ventromedial frontal cortices, together with the amygdala, would participate in a more general function of linking stimuli to emotionally valued responses, they may be notably indispensable when reasoning and making decisions about social matters.

The role of the orbitofrontal cortex in social affiliative behaviours is also of interest from a pharmacological point of view: the density of certain subtypes of serotonin receptors in the orbitofrontal cortex of monkeys correlates with the animal’s social status. Pharmacological manipulation of serotonergic neurotransmission targeted at these receptors influences social affiliative behaviour, and results in changes in social status (Raleigh et al., 1996). These findings from monkeys may offer some explanation of the changes in social behaviours which can also be observed in humans following serotonergic manipulation (for example, Prozac and ecstasy). Of special interest is the neurotransmitter serotonin (acting on specific subtypes of serotonin receptor) and the neuropeptide oxytocin, both of which appear to play a role in neurochemical systems relatively specialised for social behaviours.

B.3.5 Representing the minds of others

The mechanisms by which people represent and predict other people’s behaviour have been viewed from two different theoretical perspectives. The two main camps argue either for a ‘theory of mind’, or for a set of processes which permits ‘simulation’ of other minds. The ‘theory’-theory has been floated for some time in philosophy of mind as a possible explanation of what is commonly called ‘folk psychology’: people’s commonsense understanding of other people’s behaviour in terms of intervening mental states, such as beliefs, desires and intentions, on the basis of which people act. The other camp, however, views human ability to recognise and reason about other
people’s states of mind as an example of experience projection; in essence, people know other minds by empathy, or by simulation. It is likely that both these views have some truth to them, depending on the circumstances (Adolphs, 1999).

In the latter situation, it could be that the only way to predict what another person will do is to run in one’s own brain the processes which the other person is running in his/hers. If this possibility is taken seriously, it suggests a role for conscious experience in social cognition – to obtain information about another person’s internal mental state, it may be necessary to imagine what it would be like to be the other person via direct simulation. Simulation might find its developmental origins in infants’ ability to mimic facial expressions spontaneously (Meltzoff and Moore, 1977), and it has found some recent neurophysiological support from the finding of so-called ‘mirror neurons’, which appear to participate in simulating the actions of other individuals (Gallese and Goldman, 1999).

In humans, the theory-of-mind question was posed concretely in terms of the ability to attribute beliefs, specifically false beliefs, to other individuals. It has been shown that this ability begins to emerge around age four or possible earlier (Perner and Lang, 1999). The abilities which constitute a theory of mind have been fractionated into several distinct components, such as the ability to attribute desires, to recognise objects of shared attention, and to monitor others’ direction of gaze. All these different components appear at distinct developmental stages in humans, and there is evidence that some of them may be selectively impaired in subjects with autism (Baren-Cohen, 1995).

**B.3.6 Empathy and simulation in the somatosensory cortices**

A person must be able to represent not only his/her own body states in response to conspecific stimuli, but must also possess mechanisms for constructing detailed representations of the conspecific stimuli themselves. Social cognition should permit the construction of a mental model, a comprehensive representation of other individuals, and of what it is about those individuals that is important to know about them as social agents who have the possibility of interacting with the person. In order to answer the question of how one represents other individuals, it is useful to consider how one represents oneself. The ability to judge other people’s emotions, behavioural dispositions, beliefs and desires might draw substantially on the ability to empathise with them – that is, to create a model in one’s own mind of what the other person is feeling. It would seem that such an ability would be essential in order to adopt another person’s point of view in a
comprehensive manner, and that it would aid in the ability to predict other people’s behaviour (Adolphs, 1999).

This idea help to explain why emotion and social cognition are closely related, not only in terms of shared processing strategies, but also in terms of their neural substrates: most structures important to social cognition are also important to normal emotional functioning. The common ingredient may be what people commonly call ‘feeling’: the representation of emotional body states, either in regard to one’s own emotional reaction, or in regard to the empathy for, or simulation of, another person’s internal state.

In addition to the amygdala and ventromedial frontal cortices, which can trigger emotional responses to socially relevant stimuli, there is evidence for a third important structure which contributes directly to the ability to construct representations of other individuals. Adolphs et al. (1996a, b) found that recognition of emotions from other people’s facial expressions critically relied on the integrity of somatosensory-related cortices in the right hemisphere. In their study, somatosensory structures were particularly important in order to judge complex blends of multiple emotions in a single face. This idea proposes that subjects judge another person’s emotional state from the facial expression by reconstructing in their own brains a simulation of what the other person might be feeling. That is, subjects who are looking at pictures of facial expressions ask themselves how they would feel if they were making the facial expression shown in the stimulus (either overtly or covertly).

B.3.7 Summary

Social cognition draws upon a vast set of abilities. Some of these are quite specific to the social domain and others are more general in their application. Some classes of emotions, such as guilt, shame, embarrassment, and jealousy only make sense in a social context. Other social signals, and other types of social judgments, draw upon systems which subserve emotional processing in general, systems which permit people to build models of other individuals through simulation, and a vast network of structures which contribute to reasoning, inference and language.

Three structures have been highlighted in studies – the amygdala, ventromedial frontal cortex, and right somatosensory-related cortex. Normally, in a typical, complex, emotionally-salient situation in real life, all three component structures will operate in parallel: the amygdala will provide a quick and automatic bias with respect to those aspects of the response which pertain to evaluating the
potentially threatening nature of the situation, or with respect to allocating processing resources to
those stimuli which are potentially important but ambiguous; the ventromedial frontal cortex will
associate elements of the situation with elements of previously encountered situations, and trigger a
re-enactment of the corresponding emotional state; and right somatosensory-related cortices will be
called upon to the extent that a detailed, comprehensive representation of the body state associated
with emotional or social behaviour needs to be made available. All of these components would be
important to guide social behaviour in a typical situation in real life, and all of them emphasise the
close link between emotion and social cognition. In fact, most of the neural structures known to be
important to social cognition are also important to emotion, and to associating stimuli with reward
and punishment (Adolphs, 1999).
B.4 The Underpinnings of Emotions

B.4.1 Introduction

How do emotions influence every other aspect of a person's mental life, shaping perceptions, memories, thoughts and dreams? Do people have control over their emotions or do emotions control them? Are emotions cast in neural stone by genes or taught to the brain by the environment? Can a person have unconscious emotional reactions and unconscious emotional memories? Can the emotional slate ever be wiped clean, or are emotional memories permanent? LeDoux (1998) focuses on issues about how the brain detects and responds to emotionally arousing stimuli, how emotional learning occurs and emotional memories are formed, and how conscious emotional feelings emerge from unconscious processes.

Emotions are biological functions of the nervous system. Figuring out how emotions are represented in the brain can help people understand them. This approach contrasts sharply with the more typical one in which emotions are studied as psychological states, independent of the underlying brain mechanisms. Psychological research has been extremely valuable, but an approach where emotions are studied as brain functions is far more powerful.

The brain is an enormously rich source of variables to manipulate. Studying emotion through the brain greatly expand opportunities for making new discoveries beyond what can be achieved with psychological experimentation alone. Additionally, studying the way emotion works in the brain can help to choose between alternative psychological hypotheses – there are many possible solutions to the question of how emotions might work, but the only relevant one is the solution which was put into the brain with creation.

The proper level of analysis of a psychological function is the level at which that function is represented in the brain. This leads to a conclusion that the word ‘emotion’ does not refer to something that the mind or brain really has or does. ‘Emotion’ is only a label, a convenient way of talking about aspects of the brain and the mind. Psychology often carves the mind up into functional pieces, such as perception, memory and emotion. These are useful for organising information into general areas of research but do not refer to real functions. The brain, for example, does not have a system dedicated to perception. The word ‘perception’ describes in a general way what goes on in a number of specific neural systems – humans see, hear and smell the world with their visual, auditory and olfactory systems. In a similar vein, the various classes of
emotions are mediated by separate neural systems. There is no such thing as the ‘emotion’ faculty and there is no single brain system dedicated to this function.

When these neural systems function in a human with conscious awareness, then conscious emotional feelings occur. Emotional responses, for example trembling, sweating, and heart palpitations are, for the most part, generated unconsciously. What is an emotion but a conscious feeling? For example, take away the subjective register of fear and there is not much left to a dangerous experience. But, this idea is wrong. It is necessary to elucidate not so much the conscious state of fear or the accompanying responses, but the system that detects the danger in the first place. Fear feelings and a pounding heart are both effects caused by the activity of this system, which does its job unconsciously – literally, before a person actually knows he/she is in danger. The system which detects danger is the fundamental mechanism of fear, and the behavioural, physiological and conscious manifestations are the surface responses it orchestrates. The objectively measurable emotional responses can then be used to investigate the underlying mechanism, and, at the same time, illuminate the system which is primarily responsible for the generation of the conscious feelings. Understanding emotions in the human brain is clearly an important quest, as most mental disorders are emotional disorders.

Conscious feelings, like the feeling of being afraid or angry or happy or in love or disgusted, are in one sense no different from other states of consciousness, such as the awareness that a roundish, reddish object is an apple, that a sentence just heard was spoken in a particular foreign language, or that one has just solved a previously insoluble problem in mathematics. States of consciousness occur when the system responsible for awareness becomes privy to the activity occurring in unconscious processing systems. What differs between the state of being afraid and the state of perceiving red is not the system which represents the conscious content (fear or redness) but the systems that provide the inputs to the system of awareness. There is but one mechanism of consciousness and it can be occupied by mundane facts or highly charged emotions. Emotions easily bump mundane events out of awareness, but non-emotional events (like thoughts) do not so easily displace emotions from the mental spotlight.

Emotions are things which happen to people rather than things people will to occur. Although people set up situations to modulate their emotions all the time (for example, going to movies and amusement parks, having a tasty meal, consuming alcohol and other recreational drugs), in these situations, external events are simply arranged so that the stimuli which automatically trigger emotions will be present. People have little direct control over their emotional reactions. Anyone
who has tried to fake an emotion, or who has been the recipient of a faked one, knows all too well the futility of the attempt. While conscious control over emotions is weak, emotions can flood consciousness. This is so because the wiring of the brain is such that connections from the emotional systems to the cognitive systems are stronger than connections from the cognitive systems to the emotional systems.

Once emotions occur they become powerful motivators of future behaviours. They set the course of moment-to-moment action as well as the course towards long-term achievements. However, emotions can also get people into trouble. When fear becomes anxiety, desire gives way to greed, or annoyance turns to anger, anger to hatred, friendship to envy, love to obsession, or pleasure to addiction, emotions start working against people. Mental health is maintained by emotional hygiene and mental problems, to a large extent, reflect as breakdown of emotional order. Emotions can have both useful and pathological consequences.

The study of emotion has been ignored by the field of cognitive science, the major scientific enterprise currently concerned with the nature of the mind. Cognitive science treats minds like computers and has traditionally been more interested in how people and machines solve logical problems or play chess than in why people are sometimes happy and sometimes sad. This shortcoming has been corrected in an unfortunate way – by redefining emotions as cold cognitive processes, stripping them of their passionate qualities. At the same time though, cognitive science has been very successful and has provided a framework that, when appropriately applied, provides an immensely valuable approach for pursuing the emotional as well as the cognitive mind. One of the major conclusions about cognition and emotion that comes from this approach is that both seem to operate unconsciously, with only the outcome of cognitive or emotional processing entering awareness and occupying people’s conscious minds and only in some instances.

B.4.2 Cognition, emotion and the computer metaphor

The human brain contains about 10 billion neurons which are wired together in enormously complex ways. Although the electrical sparks within and chemical exchanges between these cells accomplish some amazing and perplexing things, the creation of emotions stands out as one of their most amazing and perplexing feats. Emotions are at the same time obvious and mysterious. They are states of the brain people know best and remember with the greatest clarity. Yet, sometimes people do not know where they come from. Emotions can change slowly or suddenly, and their causes can be evident or opaque.
Since the time of the ancient Greeks, humans have found it compelling to separate reason from passion, thinking from feeling, cognition from emotion. These contrasting aspects of the soul, as the Greeks liked to call the mind, have in fact often been viewed as waging an inner battle for the control of the human psyche. Plato, for example, said that passions, desires, and fears make it impossible for a person to think (Flew, 1964). For him, emotions were like wild horses that have to be reined in by the intellect, which he thought of as a charioteer.

Given this long tradition of separation of passion and reason, it is not too surprising that a field currently exists to study rationality, or cognition, on its own, independent of emotions. This field, known as cognitive science, tries to understand how people come to know their world and use their knowledge to live in it. It asks how people recognise a certain pattern of visual stimulation falling on the retina as a particular object, say an apple, or determine the apple's colour, or judge which of two apples is bigger.

Cognitive science emerged recently, around the middle of the 20th century, and is often described as the 'new science of mind' (Gardner, 1987). However, cognitive science is a science of only a part of the mind, the part having to do with thinking, reasoning and intellect. It leaves out emotions.

Throughout much of the first half of the twentieth century, psychology was dominated by behaviourists, who believed that the subjective inner states of mind, like perceptions, memories and emotions, are not appropriate topics for psychology (Skinner, 1938). In their view, psychology should not be the study of consciousness, but instead should be the study of observable facts – objectively measurable behaviours. Being subjective and unobservable (except by introspection), consciousness could not, in the behaviourist’s mind, be examined scientifically.

By the mid-20th century, the behaviourist stronghold on psychology began to weaken (Gardner, 1987). Electronic computers had been developed, and engineers, mathematicians, philosophers and psychologists saw similarities in the way computers process information and the way minds work. Computer operations became a metaphor for mental functions, and the field of Artificial Intelligence (AI), which seeks to model the human mind using computer simulations, was born. Anyone who bought into the notion of the mind as an information-processing device came to be known as a cognitive scientist. Cognitive science caused a revolution in psychology, dethroning the behaviourists and restoring the inner states of mind as appropriate topics.
One of the most important conceptual developments in the establishment of cognitive science was a philosophical position known as functionalism, which holds that intelligent functions carried out by different machines reflect the same underlying process (Putnam, 1966). For example, a computer and a person can both add $2 + 5$ and come up with $7$. The fact that both achieve the same answer cannot be explained by the use of similar hardware – brains are made of biological matter and computers of electronic parts. The similar outcome must be due to a similar process that occurs at a functional level. In spite of the fact that the hardware in the machines is vastly different, the software or programme that each executes may be the same. Functionalism thus holds that the mind is to the brain as a computer programme is to the computer hardware. Cognitive scientists, under the functionalist banner, have been allowed to pursue the functional organisation of the mind without reference to the hardware that generates the functional states. Regardless of whether they do experiments on humans or use computer simulations of the human mind, many cognitive scientists today are functionalists.

Cognitive scientists tend to think of the mind in terms of unconscious processes rather than conscious contents. In leaving out consciousness, cognitive science left behind those conscious states called emotions. Rooted in the idea of mind as an information-processing device, cognitive science has been geared towards understanding the functional organisation and processes that underlie and give rise to mental events, and much less towards understanding the nature of consciousness and its subjective contents. As Lashley (1950) pointed out, conscious content comes from processing, and humans are never consciously aware of the processing itself but only of the outcome. These mental processes are the essence of cognitive science. Cognitive scientists sometimes speak of consciousness as the end result of processing, but are usually far more interested in the underlying processes than in the contents of consciousness that occur during and as a result of the processing. This emphasis on unconscious processes as opposed to conscious content underlies much work in cognitive science (Neisser, 1976).

Kihlstrom (1987) coined the term ‘cognitive unconscious’ to describe the subterranean processes that have been the main preoccupation of cognitive science. These processes span many levels of mental complexity, all the way from the routine analysis of the physical features of stimuli by sensory systems, to remembrance of past events, to speaking grammatically, to imagining things which are not present, to decision-making and beyond.

The first level of analysis of any external stimulus by the nervous system involves the physical properties of the stimulus. These low-level processes occur without awareness (Ullman, 1984).
The brain has, for example, mechanisms for computing the shape, colour, location and movement of objects one sees, and the loudness, pitch and location of sounds one hears. On the basis of its analysis of physical features of stimuli, the brain begins to construct meaning. In order to know that the object the person is looking at is an apple, the physical features of the stimulus have to find their way into the person’s long-term memory banks. Once there, the stimulus information is matched up with stored information about similar objects and is classified as an apple, allowing the person to ‘know’ that he/she is looking at an apple and perhaps even leading the person to remember past experiences he/she had which involved apples. The end result is the creation of conscious memories (conscious contents), but through processes which the person has little conscious access to. Just because the brain can do something does not mean that ‘the person’ knows how it did it.

Speech, consciousness’ behavioural tool, is also the product of unconscious processes (Pinker, 1994). People do not consciously plan the grammatical structure of the sentences they utter. There simply is not enough time. People are not all good orators, but they usually say things which make sense linguistically. Speaking is one of the many things which the cognitive unconscious takes care of.

The cognitive unconscious also extends to complex judgments about the mental origins of beliefs and actions. For their study, Nisbett and Wilson (1977) created a number of carefully structured experimental situations in which people were required to do things and then say why they did what they have done. In one study, they lined up several pairs of stockings on a table. Female subjects were then allowed to examine the stockings and to choose which one they liked best. When the women were questioned, they had all sorts of answers about the texture and sheerness of the stockings that justified their choices. However, unbeknownst to them the stockings were identical. The subjects believed that they had decided on the basis of their internal judgments about the quality of the stockings. In this and a host of other studies, Nisbett and Wilson showed that people are often mistaken about the internal causes of their actions and feelings. Although the subjects always gave reasons, the reasons come not from privileged access to the processes that underlay their decisions, but from social conventions, or ideas about the way things normally work in such situations, or just plain guesses. Accurate introspective reports, Nisbett and Wilson say, often occur in life because the stimuli involved in causing the behaviour or the belief are salient and plausible causes of these. However, when salient and plausible stimuli are not available, people make up reasons and believe in them. In other words, the inner workings of important aspects of the mind, including people’s own understanding of why they do what they do, are not necessarily knowable to
the conscious self. One has to be very careful when using verbal reports based on introspective analyses of one’s own mind as scientific data.

Gazzaniga and LeDoux (1978) were engaged in studies of split-brain patients which led them to a similar conclusion. Like Nisbett and Wilson’s subjects, the patient was attributing explanations to situations as if he had introspective insight into the cause of the behaviour when in fact he did not. It was concluded that people normally do all sorts of things for reasons they are not consciously aware of (because the behaviour is produced by brain systems that operate unconsciously) and that one of the main functions of consciousness is to keep people’s lives tied together into a coherent story, a self-concept. It does this by generating explanations of behaviour on the basis of self-image, memories of the past, expectations of the future, the present social situation and the physical environment in which the behaviour is produced (Gazzaniga, 1988).

![Figure B.8](http://scholar.sun.ac.za)

**Figure B.8** A contemporary map of some cortical functions (LeDoux, 1998)

It seems clear that much of mental life occurs outside of conscious awareness. People can have introspective access to the outcome of processing (in the form of conscious content), but not all processing gives rise to conscious content. Stimulus processing which does not reach awareness, in the form of conscious content can nevertheless be stored implicitly or unconsciously and have important influences on thought and behaviour at some later time (Bowers and Meichenbaim, 1984;
Bargh, 1992). Furthermore, information can be simultaneously processed separately by systems that do and do not give rise to conscious content, leading to the conscious representation in some and the unconscious representation in other systems (Figure B.8).

That much of the processing involved in these functions occurs unconsciously has allowed cognitive science a luxury that earlier forms of mentalism did not have – the field could get on with the business of studying the mind without having to first solve the problem of consciousness (Churchland, 1984). In fact, it is probably true that consciousness will only be understood by studying the unconscious processes which make it possible.

What is it about emotion that has compelled cognitive scientists to separate it out from attention, perception, memory and other bona fide cognitive processes? Why was emotion banned from the rehabilitation of the mind that took place in psychology’s cognitive revolution? When the computer metaphor was developed, it was seen as more applicable to logical reasoning processes than to so-called illogical emotions. However, cognition is not as logical as it was once thought and emotions are not always so illogical.

AI researchers realised that knowledge was needed in problem-solving machines – problem solvers with impeccable logic but without facts did not get too far (Gardner, 1987). Knowledge had to support logic in these models. It is now believed that thinking does not normally involve the pure reasoned rules of logic. This has been demonstrated in research by Johnson-Laird (1988). He found that quite often people draw logically invalid conclusions, suggesting that if the human mind is a formal logic machine, it is a poor one. People are rational, according to Johnson-Laird, but they just do not achieve their rationality by following formal laws of logic. People use what Johnson-Laird calls mental models, hypothetical examples drawn from past experiences in real life or from imagined situations. Other studies by Kahneman et al. (1982) led to a similar view, but from a different angle. They showed that people use their implicit understanding of the way the world works, often relying on educated guesswork rather than formal principles of logic, to solve the problems that they face in their daily lives. However, Frank (1988) argues that decision-making is often not rational at all in that many actions, purposely taken with full knowledge of their consequences, are irrational. If people did not perform them, they would be better off and they know it. If cognition is not just logic, and is sometimes illogical, then emotion might not be as far afield from cognition as it was initially thought.
Success in life, according to Goleman (1995), depends on a high EQ (emotional quotient) as much or more than a high IQ. It is true that derailed emotions can lead to irrational and even pathological consequences, but emotions themselves are not necessarily irrational. While early AI programmes were most successful at modelling logical processes, more recent models have gone far beyond this truly artificial approach and some try to model aspects of emotions. Some programmes use emotional ‘scripts’ or ‘schemas’ as aids to decision making and action, others try to simulate the processes through which people evaluate or appraise the emotional meanings of stimuli, and still others attempt to make use of the understanding of the emotional brain in order to model how emotions are processed (Dyer, 1987; Scherer, 1993b; Frijda and Swagerman, 1987; Sloman, 1987; Armony et al., 1995). The logical/illogical or rational/irrational distinction is not very sharp when it comes to separating emotion and cognition, and certainly not a clean way of defining what a science of mind should be about.

Subjective emotional states, like all other states of consciousness, are best viewed as the end result of information processing occurring unconsciously. Just as one can study how the brain processes information unconsciously in perceiving visual stimuli and using visual information to guide behaviour, one can study how the brain processes the emotional significance of stimuli unconsciously and uses this information to control behaviours appropriate to the emotional meaning of the stimuli.

Some cognitive scientists have recognised that emotion is important. Simon (1967), for example, argues that cognitive models need to account for emotions in order to approximate real minds. These suggestions by leading cognitive scientists have finally begun to have an impact – more and more cognitive scientists are getting interested in emotions. The problem is, instead of heating up cognition, this effort has turned emotion cold – in cognitive models, emotions, filled with and explained by thoughts, have been stripped of passion. Minds have thoughts as well as emotions and the study of either without the other will never be fully satisfying. To call the study of cognition and emotion cognitive science is to do it a disservice.

Reacting to the functionalism that the mind can be modelled independent of knowledge of how the brain works, Churchland and Sejnowski (1990) have argued that nature is more ingenious than humans are. Humanity stands to miss all that power and ingenuity unless neurobiological plausibility is attended to. The point is, it has already been done in nature, so why not learn how the stupendous machine, the human brain, actually works?
The functionalist conception of mind as a programme that can run on any machine (mechanical, electronic, biological) has been easy to accept, or at least tolerate, in the area of cognition. The biological machine of relevance to cognition, of course, is the brain. The idea that the brain is a cognitive computer is now commonplace. However, in emotions, unlike in cognitions, the brain does not usually function independently of the body. Many if not most emotions involve bodily responses. No such relation exists between cognitions and actions. In the case of cognitively driven responses, the response is arbitrarily linked to cognition. This is partly why cognition is so powerful – cognitions allow people to be flexible, and to choose how they will respond in a certain situation. If the biological machine of emotion, but not cognition, crucially includes the body, then the kind of machine that is needed to run emotion is different from the kind needed to run cognition. Even if the functionalist argument (that the hardware is irrelevant) could be accepted for mind as cognition (and it is not clear that it can), it would not seem to work for the emotional aspects of the mind (since the hardware does seem to make a difference when it comes to emotion).

Programming a computer to be conscious would be an essential first step towards programming it to have a full-blown emotional experience, since the feelings through which people know their emotions occur when they become conscious of the unconscious workings of emotional systems in the brain. However, even if a computer could be programmed to be conscious, it is not guaranteed that it could be programmed to have an emotion, as a computer does not have the right kind of composition, which does not come from the clever assembly of human artefacts but a specialised biological creation.

B.4.3 Appraising emotions

James (1884) conceived of an emotion in terms of a sequence of events which starts with the occurrence of an arousing stimulus and ends with a passionate feeling, a conscious emotional experience. However, from James onward an important gap was left in the causal chain leading from a stimulus to emotional responses and emotional experiences. The gap occurs between the arrival of the emotion-provoking stimulus and the resulting physiological responses and/or feelings. In James’ theory, the perception of the stimulus automatically (without conscious participation) produces the responses which provide the feedback defining the feeling. However, not all stimuli which are perceived do this. Something else has to happen, and something like appraisal is needed. The physical features of the stimulus have to be evaluated – appraised; their significance to the individual has to be determined. It is this computed significance that starts the emotion ball rolling. This is the case for all of the appraisal theories (Frijda, 1993; Scherer 1993a). The brain has to
evaluate a stimulus and decide whether that stimulus should be ignored or should lead to some reaction. Appraisal, in other words, fills the gap between stimuli and responses and between stimuli and feelings. However, like James’ theory, appraisal theories are also incomplete, as the appraisal models require that the appraisal mechanism get all involved in introspectively accessible levels of higher cognition from the start.

The inadequacy of any approach to emotion based solely or mainly on introspectively accessible aspects of the mind is apparent from the experimental studies (Zajonc, 1980; Bornstein, 1992) showing that much of emotional processing occurs (or can occur) unconsciously, as well as by the fact that people often find their emotions puzzling. Consciously appraisal processes cannot be the way, or at least not the only way, the emotional brain works. Even when people are conscious of the outcome of some emotional appraisal (for example, knowing that they dislike someone), this does not mean that they consciously understand the basis of the appraisal (knowing why they dislike the person). The conscious outcome might be based on non-verbalisable intuitions, so-called gut feelings (Damasio, 1994), rather than on some verbalisable set of propositions.

Noting that emotions can sometimes be puzzling, Rorty (1980) makes a distinction between the apparent cause of an emotion (the stimuli immediately available and consciously perceived) and the actual cause. The real cause of an emotion is not necessarily some immediately present stimuli, but instead may involve the interaction of these with a causal history stored in memory. Unnoticed events can activate memories, including emotional memories, implicitly (without awareness), and implicit and undetected meanings of consciously perceived stimuli can do the same. For example, a father who yells at his children may rationalise his outburst by saying that the children were misbehaving. However, the outburst may also be due in part to the fact that he had a bad day at the office, or even to the way his parents treated him as a child, and at the time he may not be consciously aware of these influences at all. In other words, the cause of an emotion can be very different from the reasons people use to explain the emotion to themselves or others afterwards. Appraisal theories have dealt with reasons rather than causes.

Introspective understanding of the causes of emotion states can be weak, especially when people are asked to reflect back on an episode after it is over (Ericsson and Simon, 1984). Moreover, even if they are asked right away they may still not know the actual cause. There is much more to explain about an emotion than what one can get at from retrospective consciously accessible thoughts about the situation.
Bowers (1984) makes the interesting point that if people’s understanding of the causation of thought and action were directly available to introspection, the field of psychology would not be needed. Indeed, it was the inadequacy of introspection which led to behaviourism, and the success of cognitive science as an alternative to behaviourism is due to its ability to investigate the mind without relying exclusively or mainly on introspection. Scherer (1993a) also suggests that appraisal researchers turn to brain science to try to validate mechanisms which psychologists uncover. Furthermore, LeDoux argues that one might turn to brain research to find novel mechanisms which psychologists have not thought of or to find novel interpretations of existing mechanisms.

Emotion and cognition are best thought of as separate but interacting mental functions mediated by separate but interacting brain systems:

a) The perceptual (cognitive) representation of an object and the evaluation of the significance of an object are separately processed by the brain.

b) The emotional meaning of a stimulus can begin to be appraised by the brain before the perceptual systems have fully processed the stimulus.

c) The brain mechanisms through which memories of the emotional significance of stimuli are registered, stored and retrieved are different from the mechanisms through which cognitive memories of the same stimuli are processed.

d) The systems that perform emotional appraisals are directly connected with systems involved in the control of emotional responses. Once an appraisal is made by these systems, responses occur automatically. In contrast, systems involved in cognitive processing are not so tightly coupled with response control systems. The hallmark of cognitive processing is flexibility of responses on the basis of processing. Cognition gives people choices.

e) The linkage of appraisal mechanisms with response control systems means that when the appraisal mechanism detects a significant event, the programming and often the execution of a set of appropriate responses will occur. The net result is that bodily sensations often accompany appraisals and when they do they are a part of the conscious experience of emotions. Because cognitive processing is not linked up with responses in this obligatory way, intense bodily sensations are less likely to occur in association with mere thoughts.

An emotion is not merely a collection of thoughts about situations. It is not simply reasoning. It cannot be understood by just asking people what went on in their minds when they had an emotion. Emotions are notoriously difficult to verbalise. They operate in a neural space which is not readily accessed from consciousness. Psychiatrists’ and psychologists’ offices are kept packed for this
reason. Still, much of the understanding of the way the emotional mind works has been based on studies which have used verbal stimuli as the gateway to emotions or verbal reports to measure emotions.

In contrast to consciousness and its associate, natural language, unconscious processing is the rule rather than the exception throughout the creation. Moreover, in the unconscious mental realm nonverbal processing is the standard. Given that so much work on unconscious processing (cognitive and emotional) has focused on verbal processes, the picture of the level of sophistication of unconscious processes in humans is probably highly inaccurate. The workings of human unconscious processes will only be fully understood once the use of verbal stimuli and verbal reports is phased out.

B.4.4 Nature of emotions

The defining feature of an emotion is the subjective feeling that comes with it. The reason for this is that the basic building blocks of emotions are neural systems which mediate behavioural interactions with environment, particularly behaviours which take care of fundamental problems of survival. While all animals have some version of these survival systems in their brains, feelings can only occur when a survival system is present in a brain that also has the capacity for consciousness. It seems that the brain systems underlying certain emotional behaviours have been preserved throughout many levels of different species.

Some of the differences between the basic emotions lists of different investigators have to do with the words used rather than with emotions implied by the words (Izard, 1992). For example, joy and happiness, basic emotions in different lists, are probably just different names for the same emotion. If one allows these kinds of translations, there turns out to be a good deal of overlap of the different lists: many if not most of the lists include some version of fear, anger, disgust and joy. Most of the remaining disagreement is over the fringe cases, like interest, desire and surprise. The basic emotions theorists are not as divergent as they appeared, and, at least for some emotions, the evidence for an innate, biological organisation is quite strong.

To the extent that emotional responses are different, it seems that there must be different brain systems to take care of these different kinds of functions. Lumping all of these together under the unitary concept of emotional behaviour provides a convenient way for distinguishing behaviours called emotional from those which reflect cognitive functions. However, the use of a label, like
emotional behaviour', should not necessarily lead one to assume that all of the labelled functions are mediated by one system of the brain.

At neural level, each emotional unit can be thought of as consisting of a set of inputs, an appraisal mechanism, and a set of outputs. The appraisal mechanism is programmed by nature to detect certain input or trigger stimuli which are relevant to the function of the network. These are called 'natural triggers'. However, the appraisal mechanism also has the capacity to learn about stimuli which tend to be associated with and predictive of the occurrence of natural triggers and is called 'learned triggers'.

In his research, LeDoux focuses on the fear system of the brain. The fear system is understood as well or better than other emotional systems. Understanding how this system is organised can help
to consider the manner in which other emotions are organised in the brain, and how these relate to the fear system.

The fear system is not, strictly speaking, a system that results in the experience of fear. It is a system which detects danger and produces responses which maximise the probability of surviving a dangerous situation in the most beneficial way. In other words, it is a system of defensive behaviour. Interactions between the defence system and consciousness underlie feelings of fear, but the function of the defence system is survival in the face of danger. Feelings of fear are a by-product of two neural systems: one that mediates defensive behaviour and one that creates consciousness. Either one alone is not enough to produce subjective fear.

Not only is fear conditioning quick, it is also very long lasting. There is little forgetting when it comes to conditioned fear and the passing of time is not enough to get rid of it (Campeau et al., 1990). Nevertheless, repeated exposure to the conditioned stimulus (or learned trigger) in the absence of the unconditioned stimulus (or natural trigger) can lead to 'extinction'. However, extinguished conditioned fear responses can be 'reinstated' by exposure to the unconditioned stimulus or some other stressful event (Campbell and Jaynes, 1966). Spontaneous recovery, renewal, and reinstatement suggest that extinction does not eliminate the memory that the conditioned stimulus was once associated with danger, but instead reduces the likelihood that the conditioned stimulus (learned trigger) will elicit the fear response.

It had long been thought that monkeys have an inherited fear of snakes, so that the first time a monkey saw a snake it would act afraid and protect itself. However, Mineka et al. (1984) showed that laboratory-reared monkeys are in fact not afraid on the first exposure to a snake. If the young monkey is shown the snake when separated from its mother, it does not act afraid. It appears that the infant learns to be afraid of the snakes by seeing its mother acting afraid. In the presence of their mothers, young monkeys do not learn about non-frightening things (like pressing a bar for food) in this way, suggesting that there is something special about biologically relevant stimuli which make them susceptible to rapid and potent observational learning. Humans learn many things by observing others in social situations and it has been proposed that anxiety, especially pathological anxiety, is sometimes or even often learned by social observation (Bandura, 1969).

There is no denying that genes make humans different from one another and explain at least part of the variability in the way different people act in dangerous and other situations. However, one has to be very careful in interpreting differences in behaviour between different people. Dawkins
(1982) states that if a person is homozygous for a gene $G$, nothing save mutation can prevent passing $G$ on to all his/her children. So much is inexorable, but whether or not the person, or his/her children, show the phenotypic effect normally associated with possession of $G$ may depend very much on how they are brought up, what diet or education they experience and what other genes they happen to possess.

It can be concluded that genes give the raw materials out of which to build emotions. They specify the kind of nervous system humans will have, the kinds of mental processes in which it can engage and the kinds of bodily functions it can control. However, the exact way humans act, think and feel in a particular situation is determined by many other factors and is not predestined in genes. Some, if not many, emotions do have a biological basis, but social, which is to say cognitive, factors are also crucially important. Nature and nurture are partners in humans’ emotional life and the difficulty is to figure what their unique contributions are.

### B.4.5 Anatomy and physiology of the fear system

Only a few neural links separate any particular set of neurons in the brain from most others. However, the criticism that figuring out connections between brain areas is a waste of time, since information reaching one area can eventually influence many, is misplaced. By way of a small number of acquaintances, people are each potentially connected to everyone else in the world (Guare, 1990). Yet, people only get around to meeting a small subset of the earth’s population in their lives. Communication between people, like the flow of information between neurons, is selective. Although the trillions of connections made by the billions of neurons in the brain may seem to constitute a hopelessly complex web of relations, very systematic patterns of interactions exist between neurons in various brain areas.

The amygdala is a small region in the forebrain, named by the early anatomists for its almond shape. It is one of the areas of the limbic system (a useful but imprecise anatomical shorthand for the forebrain areas located between the hypothalamus and the neocortex) and had long been thought of as being important for various forms of emotional behaviour. The amygdala is composed of about a dozen subregions, but it is clear from experiments that not all are involved in fear conditioning. However, the discovery of a pathway which could transmit information directly to the amygdala from the thalamus suggested how a conditioned fear stimulus could elicit fear responses without the aid of the cortex. The direct thalamic input to the amygdala simply allows the cortex to be bypassed. The fact that emotional learning can be mediated by pathways which
bypass the neocortex is intriguing, for it suggests that emotional responses can occur without the involvement of the higher processing systems of the brain, systems believed to be involved in thinking, reasoning and consciousness.

The lateral nucleus (LA) is the gateway into the amygdala. Stimuli from the outside world are transmitted to LA, which then processes the stimuli and distributes the results to other regions of the amygdala, including the basal (B), accessory basal (AB) and central nuclei (CE) (Figure B.10). The central nucleus is the main connection with areas which control emotional responses. Different outputs of the central nucleus regulate the expression of different responses.

![Figure B.10 Organisation of information-processing pathways in the amygdala (LeDoux, 1998)](image)

Neurons in the area of the thalamus which projects to the primary auditory (and visual) cortex are narrowly tuned – they are very particular about what they will respond to. However, cells in the thalamic areas which project to the amygdala respond to a much wider range of stimuli and are said to be broadly tuned.

Although the thalamic system cannot make fine distinctions, it has an important advantage over the cortical input pathway to the amygdala in terms of time (Figure B.11). In the brain of a rat, it takes about 12 msec for an acoustic stimulus to reach the amygdala through the thalamic pathway and almost twice as long through the cortical pathway. The thalamic pathway is thus faster. However, because the direct pathway bypasses the cortex, it is unable to benefit from cortical processing. As a result, it can only provide the amygdala with a crude representation of the stimulus. It is thus a quick and dirty processing pathway. The direct pathway allows a person to begin to respond to potentially dangerous stimuli before he/she fully knows what the stimulus is. This can be very useful in dangerous situations. However, its utility requires that the cortical pathway be able to
override the direct pathway. It is possible that the direct pathway is responsible for the control of emotional responses unbeknownst to a person. This may occur in all humans some of the time, but may be a predominant mode of functioning in individuals with certain emotional disorders.

![Diagram of the low and high roads to the amygdala](image)

**Figure B.11** The low and high roads to the amygdala (LeDoux, 1998)

The responsibility of the cortex is to prevent the inappropriate response rather than to produce the appropriate one. For example, suppose walking in the woods and seeing a slender curved shape on the path. The curvature and slenderness reach the amygdala from the thalamus, whereas only the cortex distinguishes a coiled up snake from a curved stick. If it is a snake, the amygdala is ahead of the game. From the point of view of survival, it is better to respond to potentially dangerous events as if they were in fact for real than to fail to respond. The cost of treating a stick as a snake is less, in the end, than the cost of treating a snake as a stick (Figure B.12). A similar explanation can be given when driving along a highway and seeing a number of thin parallel lines running across the surface of the road. The cortex would be needed to distinguish a speed trap from among many other possibilities. However, responding immediately through braking could save the driver a speed fine.

Thus, the outline of a fear reaction system involves parallel transmission to the amygdala from the sensory thalamus and sensory cortex. The subcortical pathways provide a crude image of the external world, whereas more detailed and accurate representations come from the cortex. While the pathway from the thalamus only involves one link, several links are required to activate the amygdala by way of the cortex. Since each link adds time, the thalamic pathway is faster. Interestingly, the thalamo-amygdala and cortico-amygdala pathways converge in the lateral nucleus of the amygdala. In all likelihood, normally both pathways transmit signals to the lateral nucleus,
which appears to play a pivotal role in coordinating the sensory processes that constitute the conditioned fear stimulus. Once the information has reached the lateral nucleus it can be distributed through the internal amygdala pathways to the central nucleus, which then unleashes the full repertoire of defensive reactions.

Consider another example of a person, while walking down a street, notices someone else running towards him/her who turns out to be a mugger. For the person, the sight of someone running towards him/her becomes a conditioned fear stimulus which, if it happens again, chances are a set of standard fear responses will be set into play. However, suppose the person later finds himself/herself on the street where he/she has been mugged. Although there is no mugger, his/her body may still be going through its defence motions. The reason for this is that not only did the person get conditioned to the immediate stimulus directly associated with the trauma (the sight of the mugger running towards him/her), but also to the other stimuli that just happen to have been there. These made up the occasion or context in which the mugging took place, and like the sight of the mugger, they too were conditioned by the traumatic experience. The context has become a conditioned stimulus. During conditioning, the person is paying attention to the most obvious stimulus, but the other stimuli are bought for the same purchase price.
A context is not a particular stimulus but a collection of many. For some time it has been thought that the integration of individual stimuli into a context that no longer contains the individual elements is a function of the hippocampus (O'Keefe and Nadel, 1978; Eichenbaum and Otto, 1992). Unlike the amygdala, the hippocampus does not get information from brain regions that process individual sensory stimuli, like lights and tones (Amaral, 1987). Instead, the sights and sounds of a place are pooled together before reaching the hippocampus. This brain region then creates a representation of the context which contains not individual stimuli but relations between stimuli (Nadel and Willner, 1980; Sutherland and Rudy, 1989).

The amygdala is like the hub of a wheel. It receives low-level inputs from sensory-specific regions of the thalamus, higher level information from sensory-specific cortex and still higher level (sensory independent) information about the general situation from the hippocampal formation. Through such connections, the amygdala is able to process the emotional significance of individual stimuli as well as complex situations. The amygdala is, in essence, involved in the appraisal of emotional meaning and where stimuli do their triggering (Figure B.13).

It is not unreasonable to suggest that by knowing what the different inputs to the amygdala are, and having some idea of what function those areas play in cognition, one can get some reasonable hypotheses about what kinds of cognitive representations can arouse fear responses. By the same token, if it is known how the brain achieves some cognitive function, and it can be determined how the brain regions involved in that function are connected with the amygdala, one can explain in a plausible way how fear might be aroused by that kind of cognition. Thus, by knowing which cortical areas project to the amygdala, and knowing the functions in which those areas participate.
one can make predictions about how those functions might contribute to fear reactions. Anatomy can, in other words, illuminate psychology.

B.4.6 Emotions and memory systems of the brain

It is now known that there are multiple memory systems in the brain, each devoted to different memory functions. Different kinds of memory, like different kinds of emotions and different kinds of sensations, come out of different brain systems. Two different memory systems are used by the brain to form memories about emotional experiences (Figure B.14). The temporal lobe memory system is involved in forming declarative or explicit memories of experiences available for conscious recollection at some later time. The second system operates outside of consciousness and controls behaviour without explicit awareness of the past learning.

![Diagram](image)

**Figure B.14** Brain systems of emotional memory and memory of emotion (LeDoux, 1998)

Most researchers in the field agree about the broad outline of how the temporal lobe memory system works (Zola-Morgan and Squire, 1993). Sensory processing areas of the cortex receive inputs about external events and create perceptual representations of the stimuli. These representations are then shuttled to the surrounding cortical regions, which, in turn, send further processed representations to the hippocampus. The hippocampus then communicates back with the surrounding regions, which communicate with the neocortex. The maintenance of the memory over the short run (a few years) requires that the temporal lobe memory system be intact, either because components of this system store the memory trace or because the trace is maintained by interactions between the temporal lobe system and the neocortex. Gradually, over years, the hippocampus relinquishes its control over the memory to the neocortex essentially through long-
term potentiation (LTP), where the memory appears to remain as long as it is a memory, which may be a lifetime.

The major link between the hippocampus and neocortex is the transition cortex. Each of the major sensory processing systems of the neocortex gives rise to projections to areas which are transitional between the neocortex and hippocampus (that is, the perirhinal and parahippocampal areas). These areas send their outputs to the entorhinal cortex. Once a cortical sensory system has done all that it can do with a stimulus, for example a sight or a sound, it ships the information to the transition region, where the different sensory modalities can be mixed together. The entorhinal cortex then provides the main source of inputs to the hippocampal formation. This means that in the transition circuits, representations of the world are formed which are no longer just visual, auditory or olfactory, but which include all of these at once. One begins to leave the purely perceptual and enter the conceptual domain of the brain. The transition region then sends these conceptual representations to the hippocampus, where even more complex representations are created. The hippocampus projects back to the neocortex by way of the same pathways through the transitional cortex. Findings suggested that the learning and remembrance of manual skills might be mediated by some system other than the temporal lobe system.

In terms of fear conditioning, another memory system forms implicit or non-declarative memories about dangerous or otherwise threatening situations. Memories of this type are created through mechanisms of fear conditioning. The learning which occurs does not depend on conscious awareness and, once the learning has taken place, the stimulus does not have to be consciously perceived in order to elicit the conditioned emotional responses. Unconscious fear (and probably other emotional) memories established through the amygdala appear to be indelibly burned into the brain. Implicit, fear-conditioned memory can be called ‘emotional memory’ and explicit declarative memory a ‘memory of an emotion’.

Cohen and Squire (1984) put all of the findings together and propose that damage to the temporal lobe memory system interferes with the ability to consciously recollect, but leaves intact the ability to learn certain skills. They call these two processes declarative and procedural memory. They point out that explicit, declarative memory is mediated by a single memory system, the temporal lobe memory system, but that there are multiple implicit or procedural memory systems. Thus, the brain system which mediates priming is different from the systems involved in skill learning or classical conditioning. Further, different forms of classical conditioning are also mediated by
different neural systems — eye-blink conditioning by brain stem circuits and fear conditioning by the amygdala and its connections.

There is a place, though, where explicit memories of emotional experiences and implicit emotional memories meet — in working memory and its creation of immediate conscious experience. The co-representation in awareness of the conscious memory and the current emotional arousal give an emotional flavouring to the conscious memory. Actually, these two events (the past and the present arousal) are seamlessly fused as a unified conscious experience of the moment. This unified experience of the past memory and the arousal can then potentially be converted into a new explicit long-term memory, a memory that will include the fact that one was emotionally aroused last time the incident was remembered.

It is known from personal experience that conscious memories can make people tense and anxious and one needs to account for this as well. All that is needed for this to occur is a set of connections from the explicit memory system to the amygdala. There are in fact abundant connections from the hippocampus and the transition regions, as well as many other areas of the cortex, to the amygdala.

The idea of separate systems devoted to forming implicit emotional memories and explicit memories of emotions is relevant for understanding infantile amnesia, the inability to remember experiences from early childhood, roughly before the age of three. Jacobs and Nadel (1985) were interested in the way that early trauma, though not remembered, might have lasting, detrimental influences on mental life. They proposed that the system that forms unconscious memories of traumatic events might mature before the hippocampus. The amygdala appears to be functionally mature before the hippocampus. The separate function and differential maturation of the amygdala has important implications for understanding psychopathological conditions.

Psychologists are also aware of a phenomenon described as a ‘flashbulb memory’. Such a memory is made especially crisp and clear due to its emotional implications, for example people being able to remember exactly what they were doing when they heard that President Kennedy had been shot. Recent findings by McGaugh et al. (1995) and Gold (1992), together with the idea of separate systems for detecting the emotional implications of a situation in explicit memory, help to understand the biological basis of flashbulb memories. The role of peripheral hormones, like adrenaline, in the solidification of memory processes suggests that if adrenaline is released naturally (from the adrenal gland following activation of the autonomic nervous system by the amygdala) in some situation, that experience will be remembered especially well. Since emotional arousal
usually results in the release of adrenaline, it might be expected that the explicit conscious memory of emotional situations would be stronger that the explicit memory of non-emotional situations. Results also show that an adrenaline blockade prevents the memory-enhancing effects of emotional arousal and can help to spare rescue workers and soldiers in battle some anguish later, after been traumatised by horrific scenes they have witnessed. However, conditions (for example the intensity and duration of emotional trauma) may lead to the loss of memory as opposed to the facilitation of memory.

Explicit memories are very closely related to what is attended to during the experience. At the same time, implicit emotional memories may capture aspects of experiences which escape attention and awareness. Explicit memories, regardless of their emotional implications, are not carbon copies of the experiences that created them. They are reconstructions at the time of recall and the state of the brain at the time of recall can influence the way in which the withdrawn memory is remembered. As Bartlett (1932) demonstrated, explicit memories involve simplifications, additions, elaborations and rationalisations of learning experiences, as well as omissions of elements of the initial learning. The memory, in short, occurs in the context of what is called a cognitive schema, which includes the expectations and biases of the person doing the remembering (Erdelyi, 1985).

Learning which take place in one situation or state is generally remembered best when the one is in the same situation or state (Bower, 1992). The more cues, which were present during learning, which are also present during remembering, the more likely it is that the memory will occur. Co-activation of implicit emotional memory may help the explicit system during remembering.

Hippocampal circuits, with their massive neocortical interconnections, are well suited for establishing complex memories in which many events are bound together in space and time. The purpose of these circuits, according to Eichenbaum and Otto (1992), is to provide representational flexibility. No particular response is associated with these kinds of memories – they can be used in many different ways in many different kinds of situations. In contrast, the amygdala is more suited as a triggering device for the execution of survival reactions. Stimulus situations are rigidly coupled to specific kinds of responses through the learning and memory functions of this brain region. It is wired to pre-empt the need for thinking about what to do.
If looked at microscopically, which is to say molecularly, implicit (unconscious) emotional memory and explicit (conscious) memory of emotion may be indistinguishable. However, at the level of neural systems and their functions, these are clearly unique operations of the brain.

B.4.7 Stress and loss of memory

During a traumatic learning situation, conscious memories are laid down by a system involving the hippocampus and related cortical areas and unconscious memories established by fear conditioning mechanisms operating through an amygdala-based system. These two systems operate in parallel and store different kinds of information relevant to the experience. When stimuli which were present during the initial trauma are later encountered, each system can potentially retrieve its memories. In the case of the amygdala system, retrieval results in expression of bodily responses which prepare for danger, and in the case of the hippocampal system, conscious remembrances occur.

When the amygdala detects danger, it sends messages to the hypothalamus, which in turn sends messages to the pituitary gland, and the result is the release of a hormone called ACTH. ACTH flows through the blood to the adrenal gland to cause the release of steroid hormone. In addition to reaching target sites in the body, the steroid hormone flows through the blood into the brain, where it binds to receptors in the hippocampus, amygdala, prefrontal cortex and other regions. Because adrenal and pituitary secretions are reliably elicited by stressful events, they are called stress hormones.

It has been recognised for some time that the hippocampal steroid receptors are part of a control system that helps regulate how much adrenal steroid hormone is released (Jacobson and Sapolsky, 1991). When the hormone binds to receptors in the hippocampus, messages are sent to the hypothalamus to tell it to inform the pituitary and adrenal glands to slow down the release. In the face of stress, the amygdala keeps saying ‘release’ and the hippocampus keeps saying ‘slow down.’ If stress persists too long, the hippocampus begins to falter in its ability to control the release of the stress hormones and to perform its routine functions.

McEwen (1992) has shown that severe but temporary stress can result in a shrivelling up of dendrites in the hippocampus of an animal. Dendrites are the parts of neurons which receive incoming inputs and which are responsible, in large part, for the initial phases of long-term potentiation and memory formation (Koch et al., 1992). McEwen has also shown that if the stress
is discontinued, these changes are reversible. However, with prolonged stress, irreversible changes take place. Cells in the hippocampus actually begin to degenerate. When this happens, the memory loss is permanent. Recent studies have shown that the human hippocampus is vulnerable to stress (Bremner et al, 1995).

Stress does not interfere with the workings of the amygdala. It is thus completely possible that one might have poor conscious memory of a traumatic experience, but at the same time form very powerful implicit, unconscious emotional memories through amygdala-mediated fear conditioning.

The prefrontal cortex, like the hippocampus, may also be altered by stress. Recent research has shown that the prefrontal cortex, like the hippocampus, offers a counterforce that keeps too much of the stress hormones from being released (Diorio et al., 1993). Since prolonged stress results in a breakdown in this negative feedback control function, it may be the case that both the prefrontal cortex and hippocampus are adversely affected. A stress-induced shutdown of the prefrontal cortex might release the brakes on the amygdala, making new learning stronger and more resistant to extinction and possibly allowing previously extinguished conditioned fears to be expressed anew. Extinction therapy regulates or prevents the expression of conditioned fear responses, but does not erase the implicit memories which underlie these responses (Christianson, 1992). Extinction, in other words, involves the cortical control over the output of the amygdala rather than a wiping clean of the memory slate of the amygdala.

B.4.8 Model of emotions

LeDoux’s idea about the nature of conscious emotional experiences and feelings constitutes that a subjective emotional experience, like the feeling of being afraid, results when one becomes consciously aware that an emotion system of the brain, like the defence system, is active. Emotional experience is not a problem about emotion. Instead, it is a problem about how conscious experiences occur.

Emotions researchers certainly have a lot to contribute to the study of consciousness, but figuring out consciousness is not their job, or at least theirs alone. Although this may seem obvious, the study of emotion has been so focused on the problem of emotional consciousness that the basic underlying emotional mechanisms have often been given short shrift. By treating emotions as unconscious processes which can sometimes give rise to conscious content, the burden of the mind-body problem is lifted from the shoulders of emotion researchers and allow them to continue with
the problem of figuring out the unconscious emotional functioning of the brain. In order to understand what an emotion is and how particular emotional feelings come about one needs to understand the way the specialised emotion systems operate and determine how their activity gets represented in working memory. Feelings come about when the activity of specialised emotion systems gets represented in the system which gives rise to consciousness.

The immediate present involves memory – what one knows about the present moment is basically what is in one’s working memory. For example, Kosslyn and Koenig (1992) argue that to be aware of something, the something must be in working memory. Kihlstrom (1987) proposes that a link must be made between the mental representation of an event and a mental representation of the ‘self’ as the agent or experiencer in order to be conscious of that event.

The conscious and unconscious aspects of thought are described in terms of serial and parallel functions. Consciousness seems to function serially, whereas the unconscious mind, being composed of many different systems, seems to work more or less in parallel. The conscious processor works at the symbolic level, which yields introspectively accessible content, but the parallel processors work sub-symbolically and their operations are not directly accessible from consciousness. And since not all sub-symbolic processors necessarily feed into the consciousness processor, some sub-symbolic processing remains inaccessible. Working memory is the limited-capacity serial processor which creates and manipulates symbolic representations. In other words, working memory is likely to be an important, and possibly an essential, aspect of consciousness. It is in fact likely to be the platform on which a conscious experience stands. But consciousness, especially its phenomenal or subjective nature, is not completely explained by the computational processes which underlie working memory, at least not in a way that anyone presently comprehends (Block, 1995).

The workspace in the brain, called working memory, is a temporary storage mechanism which allows several pieces of information to be held in mind at the same time and compared, contrasted and otherwise interrelated (Baddeley, 1982). The term working memory implies not just a temporary storage system, but an active processing mechanism used in thinking and reasoning. However, working memory is not a pure product of the here and now. It also depends on what one knows and what kinds of experiences one has had in the past. In other words, it depends on long-term memory.
The lateral prefrontal cortex is believed to exist only in primates and is considerably larger in humans than in other primates (Preuss, 1995). It is not surprising that one of the most sophisticated cognitive functions of the brain should involve this region. The lateral prefrontal cortex is ideally suited to perform general-purpose working memory functions (Figure B.15). It has connections with the various sensory systems (for example the visual and auditory systems) and other neocortical systems which perform specialised temporary storage functions (for example spatial and verbal storage) and is also connected with the hippocampus and other cortical areas involved in long-term memory (Fuster, 1989; Uylings and Van Eden, 1990). In addition, it has connections with areas of the cortex involved in movement control, allowing decisions made by the executive to be turned into voluntarily performed actions (Goldman-Rakic, 1987). Recent studies have begun to show how the lateral prefrontal cortex interacts with some of these areas. Best understood are interactions with temporary storage buffers in the visual cortex.

Signals picked up by the eyes are transmitted through the visual system to the visual thalamus, and then to the visual cortex, where a sensory representation of the scene is created and held in a short-term visual object buffer. Connections from the visual cortex to the cortical long-term memory networks activate relevant memories. By way of connections between the long-term memory networks and the working memory system, activated long-term memories are integrated with the sensory representation of the stimulus in working memory.

The parietal, temporal, and frontal regions of the neocortex are anatomically interconnected – the parietal (and temporal) area sends axons to the prefrontal region and the prefrontal region sends
axons back to the parietal (and temporal) area. These findings suggest that the parietal lobe visual area works with the lateral prefrontal cortex to maintain information about the spatial location of visual stimuli in working memory. The maintenance of visual information in working memory thus appears to crucially depend on interactions between the lateral prefrontal region and specialised areas of the visual cortex. The pathway from the specialised visual areas tells the prefrontal cortex ‘what’ is out there (from the temporal lobe) and ‘where’ it is located (from the parietal lobe) through bottom-up processing. The prefrontal cortex, by way of pathways back to the visual areas, primes the visual system to attend to those objects and spatial locations which are being processed in working memory (top-down processing). These kinds of top-down influences on sensory processing are believed to be important aspects of the executive control functions of working memory.

Imaging studies in humans have shown that another area of the frontal lobe, the anterior cingulate cortex, is also activated by working memory and related cognitive tasks (D’Esposito et al., 1995). Like the lateral prefrontal cortex, the anterior cingulate region receives inputs from the various specialised sensory buffers, and the anterior cingulate and the lateral prefrontal cortex are anatomically interconnected (Fuster, 1989). Moreover, both regions are part of what has been called the frontal lobe attentional network, a cognitive system involved in selective attention, mental resource allocation, decision-making processes and voluntary movement control (Posner, 1992).

One other area of the prefrontal cortex, the orbital region, located on the underneath side of the frontal lobe, has emerged as important as well. Damage to this region interferes with short-term memory about reward information – whether a stimulus has just led to a reward or punishment. Humans with orbital frontal damage become oblivious to social and emotional cues. This area receives inputs from sensory processing systems (including their temporary buffers) and is also intimately connected with the amygdala and the anterior cingulate region. The orbital cortex provides a link through which emotional processing by the amygdala might be related in working memory to information being processed in sensory or other regions of the neocortex.

In addition to projecting back to cortical sensory areas from which it receives inputs, the amygdala also projects to some sensory processing areas from which it does not receive inputs. For example, in order for a visual stimulus to reach the amygdala by way of the cortex, the stimulus has to go through the primary visual cortex, to a secondary region, and then to a third cortical area in the temporal lobe (which does the short-term buffering of visual object information). This third area
then projects to the amygdala. The amygdala projects back to this area, but also to the other two earlier visual processing regions. As a result, once the amygdala is activated, it is able to influence the cortical areas processing the stimuli which are activating it (Figure B.16). This might be very important in directing attention to emotionally relevant stimuli by keeping the short-term object buffer focused on the stimuli to which the amygdala is assigning significance. Although the amygdala has relatively meagre connections with the lateral prefrontal cortex, it sends rather strong connections to the anterior cingulate cortex. It also sends connections to the orbital cortex, an area in working memory which may be especially involved in working memories about rewards and punishments (Figure B.17).

![Figure B.16](Image)

Figure B.16 The influence of the amygdala on sensory areas of cortex is greater than the influence of the same areas on the amygdala (LeDoux, 1998)

In sum, connections from the amygdala to the cortex allow the defence networks of the amygdala to influence attention, perception and memory in situations where people are facing danger.

In addition to the direct influences of the amygdala on the cortex, there are a number of indirect channels through which the effects of amygdala activation can impact on the cortical processing. An important set of such connections involves the arousal systems of the brain. When someone is alert and paying attention to something important, the cortex is aroused. When drowsy and not focusing on anything, the cortex is in the unaroused state. During sleep, the cortex is also in the unaroused state, except during dream sleep when it is highly aroused. In dream sleep, in fact, the cortex is in a state of arousal which is very similar to the alert waking state, except that it has no access to external stimuli and only processes internal events (McCormick and Bal, 1994).
A number of different systems appear to contribute to arousal. Four of these systems are located in regions of the brain stem. Each has a specific chemical identity, which means the cells in each contain different neurotransmitters which are released by their axon terminals when the cells are activated, namely acetylcholine (ACh), noradrenaline, dopamine, and serotonin. A fifth group, also containing ACh, is located in the forebrain, near the amygdala. The axons of each of these cell groups terminate in widespread areas of the forebrain. In the presence of novel or otherwise significant stimuli the axon terminals release their neurotransmitters and ‘arouse’ cortical cells, making them especially receptive to incoming signals.

Arousal helps lock a person into the emotional state he/she is in. This can be very useful (one does not want to get distracted when in danger), but can also be an annoyance.

Although each of the arousal systems probably contributes to arousal in the presence of stimuli which are dangerous or warn of danger, it appears that interactions between the amygdala and the nearby ACh-containing system in the forebrain, called the nucleus basalis, is particularly important (Kapp et al., 1992; Weinberger, 1995). Damage to the amygdala or to the nucleus basalis prevents stimuli which warn of danger, like conditioned fear stimuli, from eliciting arousal. Administration of drugs which block the actions of ACh in the cortex prevents these effects on arousal of conditioned stimuli, amygdala stimulation, or nucleus basalis stimulation from occurring.

The information content provided by arousal systems is weak. The cortex is unable to discern that danger (as opposed to some other emotional condition) exists from the pattern of neural messages it receives from arousal systems. Arousal systems simply state that something important is happening. All the inputs which are vying for the attention of working memory are blocked out.
Activation of the amygdala results in the automatic activation of networks which control the expression of a variety of responses: (i) species-specific behaviours (for example freezing, fighting, and facial expressions), (ii) autonomic nervous system (ANS) responses (for example changes in blood pressure and heart rate, sweating), and (iii) hormonal responses (for example the release of stress hormones, like adrenaline and adrenal steroids, as well as other peptides, into the bloodstream). The ANS and hormonal responses can be considered together as visceral responses - responses of the internal organs and glands (the viscera). When these behavioural and visceral responses are expressed, they create signals in the body which return to the brain.

Previously, the visceral systems were thought to respond uniformly in all situations. However, recent findings suggest that the ANS, which controls the viscera, has the ability to respond selectively, so that visceral organs can be activated in different ways in different situations. For example, studies show that different emotions (anger, fear, disgust, sadness, happiness, surprise) can be distinguished to some extent on the basis of different autonomic nervous system responses (like skin temperature and heart rate) (Levenson, 1992).

Regardless of their specificity, visceral responses have relatively slow actions, too slow to be the factor determining which emotion one experiences at a given moment. However, emotional states are dynamic. For example, fear can turn into anger or disgust or relief as an emotional episode unfolds, and it is possible that visceral feedback contributes to these emotional changes over time. While arousal is non-specific and tends to lock a person into the state he/she is in when the arousal occurs, unique patterns of visceral, especially chemical, feedback have the potential for altering which brain systems are active and thus may contribute to transitions from one emotion to another within a given emotional event.

A theory by Damasio (1994), the somatic marker hypothesis, calls upon the entire pattern of somatic and visceral feedback from the body. Damasio proposes that such information underlies ‘gut feelings’ and plays a crucial role in emotional experiences and decision-making processes.

When all the interactions between the various systems are taken together, the possibilities for the generation of emotion-specific patterns of feedback are staggering. Although a person may run both to get food and to escape from danger, the feedback from the somatic and visceral responses which return to the brain will interact with different systems in these two instances. The feedback from running from danger will find the food-seeking system idle but the defence system active.
The same pattern of feedback can have unique contributions when it interacts with specific brain systems.

Studies by Ekman (1993) and by Adelman and Zajonc (1989) have shown that feedback is indeed used. For example, Ekman had subjects move certain facial muscles. Unbeknownst to the subjects, they were being made to exhibit the facial expressions characteristic of different emotions. They then had to answer some questions about their mood. It turns out that the way the subjects felt was significantly influenced by whether they had been wearing positive or negative emotion expressions. Thus, putting on a happy face may not be such a bad idea when one is feeling blue.

![Figure B.18 Some neural ingredients of a conscious emotional experience (LeDoux, 1998)](image)

For LeDoux it seems impossible that a full-blooded emotional feeling could exist without a body attached to the brain which is trying to have the feeling. Ingredients of an emotional feeling (Figure B.18) needed to turn an emotional reaction into a conscious emotional experience can be summarised as:

a) A specialised emotion system which receives sensory inputs and produces behavioural, autonomic and hormonal responses.
b) Cortical sensory buffers which hold on to information about the currently present stimuli.

c) A working memory executive which keeps track of short-term buffers, retrieves information from long-term memory and interprets the contents of the short-term buffers in terms of activated long-term memories.

d) Cortical arousal keeping conscious attention directed towards the emotional situation.

e) Bodily feedback – somatic and visceral information which returns to the brain during an act of emotional responding.

Conscious emotional feelings and conscious thoughts are in some sense very similar. Both involve the symbolic representation in working memory of sub-symbolic processes carried out by systems which function unconsciously. However, the difference between them is not due to the system concerned with consciousness. Emotional feelings and mere thoughts are generated by different sub-symbolic systems, while emotional feelings involve many more brain systems than thoughts (cognition).
B.5 Explaining Consciousness and Attention

B.5.1 Theory of consciousness (Dennett, 2001)

Overview

Theorists are converging from quite different quarters on a version of the global neuronal workspace model of consciousness, but there are residual confusions to be dissolved. In particular, it is essential to resist the temptation to see global accessibility as the cause of consciousness (as if consciousness were some other, further condition); rather, it is consciousness. A useful metaphor for keeping this elusive idea in focus is that consciousness is rather like fame in the brain. It is not a privileged medium of representation, or an added property some states have; it is the very mutual accessibility that gives some informational states the powers which come with a subject’s consciousness of that information. Like fame, consciousness is not a momentary condition, or a purely dispositional state, but rather a matter of actual influence over time.

Competition for clout

At any given time, many modular cerebral networks are active in parallel and process information in an unconscious manner. Modularity comes in degrees and kinds – these modular networks are specialist networks with limited powers of information processing. An ‘information’ becomes conscious, however, if the neural population which represents it is mobilised by top-down attentional amplification into a brain-scale state of coherent activity which involves many neurons distributed throughout the brain. There is no standard term for an event in the brain which carries information of content on some topic – near-synonyms are ‘signal’ or ‘representation’. Since there is no single organisational summit to the brain, it means only that such attentional amplification is not just modulated ‘bottom-up’ by features internal to the processing stream in which it rides, but also by sideways influences, from competitive, cooperative, collateral activities whose emergent net result is what may be lumped together and call top-down influence. In an arena of opponent processes the ‘top’ is distributed, not localised. A key point is that the accessibility of the modular (specialist) networks to each other can in principle explain the dramatic increases in cognitive competence associated with consciousness – open-mindedness which permits a conscious agent to consider anything in its purview in any way it chooses.
The long distance connectivity of these ‘workplace neurons’ can, when they are active for a minimal duration, make the information available to a variety of processes including perceptual categorisation, long-term memorisation, evaluation and intentional action. This global availability of information through the workplace is what is subjectively experience as a conscious state. One should resist the temptation to imagine some other effect that needs to build up over time, because the proposed consensual thesis is not that this global availability causes some further effect or a different sort altogether – for example igniting the glow of conscious qualia, or gaining entrance to the ‘Cartesian Theater’, or ‘phenomenality with reflexivity’ – but that it is, all by itself, a conscious state.

The basic idea is that consciousness is more like fame than television; it is not a special ‘medium of representation’ in the brain into which content-bearing events must be transduced in order to become conscious. However, the neural correlates of awareness of a given perceptual attribute are found in the very neural structure which perceptually analyses that attribute. Instead of switching media or going somewhere in order to become conscious, heretofore unconscious contents, staying right where they are, can achieve something rather like fame in competition with other fame-seeking (or just potentially fame-finding) contents. And, according to this view, that is what consciousness is. Consciousness, like fame, is not an intrinsic property, and not even just a dispositional property; it is a phenomenon which requires some actualisation of the potential – there is potential fame in the brain and there is fame in the brain. Real fame is not the cause of all the normal aftermath; it is the normal aftermath.

Researchers are inclined to comment that there may indeed be fierce competition between ‘informations’ for fame (or clout) in the brain, but that there has to be a ‘First Person’ who entertains the winners. However, one has not solved the problem of consciousness until this ‘Executive’ is itself broken down into subcomponents which are themselves clearly just unconscious underlabourers which themselves work (for example compete, and interfere) without supervision. Contrary to appearances, then, those who work on answers to this problem are not leaving consciousness out; they are explaining consciousness by leaving it behind. That is to say, the only way to explain consciousness is to move beyond consciousness, accounting for the effects consciousness has when it is achieved.

The human body is made up of some trillions of cells, each one utterly ignorant of all the things the person knows. If one is to explain the conscious Subject, one way or another the transition from clueless cells to knowing organisations of cells must be made without any magic ingredients. This
requirement presents theorists with what some see as a nasty dilemma. Proposing a theory of the knowing Subject which describes whatever it describes as like the workings of a vacant automated ‘factory’ – not a Subject in sight – one will seem to many observers to have changed the subject or missed the point. On the other hand, if the theory still has tasks for a Subject to perform, still has a need for the Subject as Witness, then although one can be falsely comforted by the sense that there is still somebody at home in the brain, one has actually postponed the task of explaining what needs explaining. To Dennett one of the most fascinating bifurcations in the intellectual world is between those to whom it is obvious that a theory which leaves out the Subject is thereby disqualified as a theory of consciousness, and those to whom it is just as obvious that any theory which does not leave out the Subject is disqualified.

Conclusion

According to Dennett, a neuroscientific theory of consciousness must be a theory of the Subject of consciousness, one which analyses the imagined central ‘Executive’ into component parts, none of which can itself be a proper Subject. The apparent properties of consciousness which only make sense as features enjoyed by the Subject must therefore also be decomposed and distributed, and this inevitably creates a pressure on the imagination of the theorist. No sooner do such properties get functionalistically analysed into complex dispositional traits distributed in space and time in the brain, than their ghosts come knocking on the door, demanding entrance disguised as qualia, or phenomenality or the imaginable difference between normal persons and zombies. One of the hardest tasks thus facing those who would explain consciousness is recognising when some feature has already be ‘explained’ (in sketch, in outline) and hence does not need to be explained again.

B.5.2 A Neural Global Workspace Model for Conscious Attention (Newman et al., 1997)

Overview

Considerable progress is being made in interdisciplinary efforts to develop a general theory of the neural correlates of consciousness. Developments of the Global Workspace (GW) theory (Baars, 1997) are examples of this progress. Integrating experimental data and models from cognitive psychology and neuroscience Newman et al. developed a neurocognitive model in which consciousness is defined as a global integration and dissemination system – nested in a large-scale, distributed array of specialised bioprocessors – which controls the allocation of the processing resources of the central nervous system (CNS). Among the various functions of this system are the
allocation of processing resources based, first, upon biological contingencies of novelty, need or potential threat and, secondly, cognitive schemas, purposes and plans. It is posited that this global control is effected via cortical ‘gating’ of the thalamic nucleus. The basic circuitry of this neural system is reasonably well understood, and can be modelled to a first approximation employing neural network principles.

Conscious percepts are characterised by unified gestalts of shape, texture, colour, location and movement, despite the fact that these contributions to perception are initially processed in parallel areas of the cortex, in both hemispheres. Moreover, conscious intentions are generally single-minded and goal-directed. Of course, conflicts can and do arise, but a central purpose of consciousness seems to be resolving such conflicts.

While such global states can be highly adaptive – indeed, are essential to explicit learning – GW theory maintains that the vast majority of cognitive tasks performed by the human brain are automatic and largely non-conscious. Consciousness generally comes in play when stimuli are assessed to be novel, threatening, or momentarily relevant to active schemas or intentions.

Generally, people are conscious of what has the highest relevance to them at that moment. This may be a momentary threat, a sudden insight, or a pleasant sensation. In relaxed moments, there may be no particular focus or intent, simply a stream of associations. Yet, while the range of awareness is immense (limited only by the most developed cognitive capacities), it is contended that the basic mechanism for the allocation of these capacities remains constant under virtually all contingencies; and the basic neural circuitry of that resource-allocation mechanism is reasonably well understood and might be modelled based upon already existing neural network simulations.

In GW theory the single homunculus is replaced by a large ‘audience of experts’. The ‘theatre of consciousness’ then becomes a workspace, with stage. Almost everyone in an audience has potential access to centre stage (although most prefer to simply observe, or exert indirect influences). The focus of conscious activity, at any moment, corresponds to the ‘work’ produces by the most active coalition of experts, or modular processors: whoever has managed to win the competition for ‘the spotlight’ (or fame). There is no fixed, super-ordinate observer. Individual modules can pay as much or as little attention as suits them, based upon their particular expertise. At any one moment, some may be dozing in their seats, others busy on stage. In this sense, the global workspace resembles more a deliberative body than a theatre audience. Each expert has a certain degree of ‘influence’, and by forming coalitions with other experts can contribute to
deciding which issues receive immediate attention and which are 'sent back to committee'. Most of the work of this deliberative body is done 'off stage' (or non-consciously). Only matters of greatest relevance in-the-moment gain access to consciousness.

While the GW is a teeming multiplicity, what is explicitly represented in consciousness is largely coherent and adaptive. The overall workspace serves as a 'global integration and dissemination system', in which all experts can participate, but only select coalitions dominate, momentarily, producing an orderly succession of global representations. The stream of consciousness arises out of the operations of the GW system – and, over time, people's sense of being a coherent 'I'. It is this unitary awareness, not any agent or homunculus, that is globally super-ordinate. Of course, such a system is prone to inefficiencies and pathological perturbations, but this is consistent with the scientific literature concerning human consciousness.

Modelling global, competitive attention

Consciousness is a dynamic process, not a static structure. Also, it is not localised to some 'brain centre', but arises out of the coordinated activities of widely distributed networks of neurons. Resource allocation is integral to these activities. The neural bases of resource allocation, or attention, have been extensively explored (LaBerge, 1995). Of course not all forms of attention are conscious. As an example from Artificial Intelligence, McClelland (1986) notes that in simulations of reading, activated modules must be 'sticky', that is interactive activation processes continue in older parts of the programmable blackboard while they are being set up in newer parts as the eye moves along. This 'stickiness' would seem to entail a type of attention. It normally proceeds quite automatically in both a reading machine and in a literate person. Only when the process is disrupted by a misspelled or unknown word, for example, does that word becomes the focus of the reader's conscious awareness. Normally, the reader are only conscious of the overall sense of the passage of text, and the images and thoughts it evokes, not particular semantic or syntactical operations. These linguistic processes became second nature to the reader. Such 'particular operations' are hardly trivial aspects of language acquisition, but they tend to be automated through experience and thus rendered unconscious.

Conscious awareness clearly involves a higher order of resource allocation, called 'global attention'. The term refers to a level of cognitive processing at which a single, coherent stream of information emerges out of the diverse activities of the CNS.
The processing load of global attention is both highly chunked and highly restricted. The non-conscious allocation of processing resources operates under no such constraints. For example, neuroscience has shown that specialised areas in the visual cortex process, in parallel, the contour, movement, colour, and spatial location of a stimulus. Yet, a person’s awareness is of a single, coherent object (and often includes tactile, auditory and associative aspects). Thus, neuroscience is faced with the ‘binding problem’ of how these multifarious representations, generated by widely separated areas, are integrated into real-time ‘objects’ of perception.

One would expect the neural mechanism for global attention to be complex, and widely distributed, which it is. But the basic circuitry can be described, to a first approximation, in terms of repeating, parallel loops of thalamo-cortico-thalamic axons, passing through a thin sheet of neurons known as the nucleus reticularis thalami (nRt). The loops are formed by long-axoned, excitatory neurons. The neurons of nRt are largely GABAergic, inhibitory neurons. Most, if not all, of the looping axons give off collaterals as they pass through nRt, while nRt neurons themselves project mainly to cells of the particular thalamic nucleus lying directly beneath them. There is an orderly topography to this array of axon collaterals and underlying thalamic nuclei. It essentially mirrors, in miniature, the modular architecture of the cortex.

Evidence for the central role of this ‘thalamo-cortical circuit’ (LaBerge, 1995) in attention and consciousness has been accumulating for decades. However, the precise connectivities and physiology of the thalamo-cortical circuit are not yet fully worked out. The ‘wagon wheel’ model (next section) represents a synthesis of both the accumulated evidence and related models.

Most attentional models are based upon conventional simulations of mechanisms such as centre-surround inhibition, or winner-take-all (WTA) competitions, among local circuits. Various researchers have described the network of nRt neurons as a mosaic, or array, of neural ‘gatelets’ acting to selectively filter the flow of sensory inputs to the cortex. The WTA dynamic seems analogous to the ‘competition’ posited by GW theory. The problem with such conventional networks is that they are poorly suited to global forms of competition, because prohibitively long-range and geometrically increasing numbers of connections would be required. Moreover, most long-range, reciprocal connections in the CNS are excitatory. Inhibitory effects tend to be local.

However, Taylor and Alavi (1993) have modelled a competitive network for global attention based upon a highly simplified version of the ‘thalamus-NRT-cortex complex’. Their model is unique, in that it takes into account the effects of dendro-dendritic interactions throughout nRt – the dendrites
of nRt cells project out tangentially and bi-directionally within the reticular sheet, as shown in Figure B.19. The addition of dendro-dendritic connections to the looping circuits provided the basis for a simple version of the global gating model which instantiates a form of competition in the spatial wavelength parameters of incoming inputs. In this version of the model, the entire nRt network oscillates with a wavelength, with the net strength given by the component of the input with the same wavelength. Global control arises when only those inputs which have special spatial wavelength oscillations are allowed through to the cortex, or are allowed to persist in those regions of the cortex strongly connected to the nRt: the thalamus-nRt system acts as a spatial Fourier filter.

![Figure B.19](image)

**Figure B.19** The wiring diagram of the model of the thalamus-nRt-cortex complex. Input \( l_i \) is sent both to the thalamic relay cell \( T_j \) and the inhibitory interneuron \( IN_j \), which also feeds to \( T_j \).

Output from \( T_j \) goes to the corresponding cortical cell \( C_j \), which returns its output to \( T_j \).

Both the axons \( T_jC_j \) and \( C_jT_j \) send axon collaterals to the corresponding nRt cell \( N_j \).

There is axonal output from \( N_j \) to \( IN_j \), as well as collaterals to neighbouring nRt cells, and dendro-dendritic synapses between nRt cells. (Taylor and Alavi, 1993)

Simulation runs demonstrated the global, wave-like properties of the competitive model. As LaBerge (1995) notes, the actual physiology of nRt gating in alert states remains unclear, but it is firmly established that nRt is the source of global oscillatory activity (at 8-13 Hz) initiating the descent into sleep. The nRt cells are known to inhibit each other, and when inhibition hyperpolarises an nRt cell sufficiently, it produces a rebound burst. In this way, a network of connected
nRt inhibitory cells can spread activity to every cell within the network, apparently without decrement in the intensity of the activity. Here then, is a plausible circuitry for a global, winner-take-all competition among the large array of specialised cortical processing areas.

Llinas et al. (1994) offer an interesting variation upon this circuitry, in which thalamo-cortical loops of the ‘non-specific’ intralaminar nuclei operate in parallel with the specific (input) loops described above. The synchronous activation of specific and non-specific loops is postulated to provide a basis for perceptual unit by which different sensory components are gathered into one global image'. Their modelling is concerned with high-frequency electroencephalographic (EEG) oscillations (and omits dendro-dendritic connections), yet appears to parallel much of what is discussed above. When the interconnectivity of these nuclei is combined with the intrinsic properties of the individual neurons, a network for resonant neuronal oscillations emerges in which specific cortico-thalamic circuits would tend to resonate at 40 Hz. According to this hypothesis, neurons at the different levels, and particularly those in the reticular nucleus, would be responsible for the synchronisation of 40-Hz oscillations in distant thalamic and cortical sites. These oscillations may be organised globally over the CNS, especially as it has been shown that neighbouring reticular cells are linked by dendrito-dendritic and intranuclear axon collaterals.

A neural model for global resource allocation

In the previous section a set of convergent models for the basic circuitry of a GW system involved in the integration and dissemination of the processing resources of the nervous system has been introduced. This ‘bare bones’ version accounts for how a global, winner-take-all competition might be mediated between various external inputs and cortical modules, to produce a single, coherent stream of information out of the diverse activities of the CNS. There remains to be explained how the thalamo-cortical circuit fits in with the second half of the working definition for the conscious system: the allocation of processing resources based, first, upon biological contingencies of novelty, need or potential threat and, secondly, cognitive schemas, purposes and plans. In keeping with the definition, it is necessary to first add a sub-cortical component for orienting to ‘novelty, need, or potential threat’, and then the much more complex aspects of cortically-mediated effects upon the system can be appreciated.

The extended version of the model is schematically illustrated in Figure B.20 as a ‘wagon wheel’ with the ventral thalamus (Th) as its ‘hub’. The reticular nucleus (nRt) corresponds to the metal sleeve fitted around the hub. The upper rim of the wheel represents the cerebral cortex.
(PFC, S1, ..., V1), and closely associated basal ganglia (BG). The lower half shows the major sensory systems and subcortical nuclei whose projections converge upon the thalamus. The outer ‘spokes’ represent the sensory pathways for vision, audition and the bodily senses. These project, in an orderly topography, to modality-specific nuclei in the thalamic ‘hub’. As they ascend towards the thalamus, these pathways give off collaterals to the midbrain reticular formation (MRF).

**Figure B.20** ‘Wagon wheel’ model of CNS systems contributing to global attention and conscious perception. (A1: primary auditory area; gc: ‘closed’ nRt gate; go: ‘open’ nRt gate; PFC: prefrontal cortex; S1: primary somatosensory area; V1: primary visual cortex) 

(Newman et al., 1997)

Scheibel (1980) reviewed three decades of experimental evidence indicating that these midbrain collaterals serve as the basis for an initial ‘spatial envelope’, or global map, of the environment surrounding the organism. Most MRF neurons appear multi-modal, responding to particular visual, somatic, and auditory stimuli, with combinations of the last two stimuli most numerous. The common receptive fields of typical bimodal cells in this array show a significant degree of congruence. For instance, a unit responding to stimulation of the hind limb will usually prove maximally sensitive to auditory stimuli originating to the rear of the organism. These twin somatic
and auditory maps retain an approximate register of the surrounding environment, and overlap the visuotopic map laid down in the superior colliculus.

More recent research has supported Scheibel’s portrayal of the superior colliculus as the visual component of what Crick and Koch (1990) termed a ‘saliency map’ for eye movements, involved in orienting the organism to biologically relevant stimuli. Subsequent findings have both confirmed Scheibel’s analysis and revealed a number of ‘top-down’ projections which modulate activities in MRF. The superficial area (of the superior colliculus) receives strong cortical inputs from V1, V2 and V3 (primary and secondary visual cortex); the deep layers in the superior colliculus receive their main cortical inputs from the posterior parietal area, from the prefrontal areas and the frontal eye fields (LaBerge, 1995). The deep layers contain a map of visual space which is stacked adjacent to maps for auditory and somatosensory spaces in a manner that cells corresponding to points in space lie along the same vertical axis.

LaBerge also describes inputs to the basal ganglia which are of particular importance because they tonically inhibit activity in the SC cells. It has long been known that the frontal eye fields, and posterior parietal area exert strong influences on eye movements and must be considered together with the superior colliculus in accounting for orienting of attention. These facts emphasise two key aspects of the ‘conscious system’ as modelled: (i) it is poly-modal, integrating not just visual, auditory and somatosensory inputs, but motor and ‘higher-order’ cortical effects; and (ii) it is extended, with input/output relations reaching from the brain stem core to association cortices. Indeed, the general term been used to describe it elsewhere is the ‘extended reticular-thalamic activation system’, or ‘ERTAS’.

The third key aspect of the system (as exemplified by the ‘wagon wheel’ model) is that it converges on the thalamus. This has already been discussed in terms of the thalamo-cortical circuit, which connects to virtually every area of the cerebral cortex.

The fact that Scheibel’s (1980) ‘spatial envelope’ projects with some topographic precision upon nRt would appear to enable it to disinhibit particular arrays of nRt gatelets, selectively enhancing the flow of sensory information to the cortex. As noted above, Llinas et al. (1994) hypothesise the perceptual unity of consciousness (binding) to be brought about by the global synchronisation of specific and non-specific circuits via nRt.
Depending on the nature of the alerting stimulus or central excitation, only that portion of the nRt will open which controls the appropriate subjacent thalamic sensory field. The nRt gate thus becomes a mosaic of gatelets, each tied to some specific receptive zone or species of input. Each is under the delicate yet opposed control of: (i) the specifically signatured sensory input and its integrated feedback from cortex (S1, ..., VI), (ii) MRF with its concern more for novelty (like danger) than for specific details of experience, and (iii) the frontal granular cortex-medial thalamic system (PFC/BG) more attuned to upper level strategies of the organism, whether based on drive mechanisms or on still more complex derivative phenomenon. Perhaps here resides the structuro-functional substrate for selective awareness, and in the delicacy and complexity of its connections, a person’s source of knowing, and of knowing that he/she knows.

It is essential to tie the operations of this thalamus-centred system more closely to those of the cortex and basal ganglia, or most of the functions routinely studied by cognitive science have no place in the model. This introduces an exponentially higher level of complexity. One of the values of GW theory, however, is that it provides a framework for understanding this complexity. First, it holds that the vast majority of cognitive functions are carried out, non-consciously, via changing arrays of specialised, modular processors. This is reflected, anatomically, in the immense number of cortico-cortical connections in the human brain, outnumbering those with subcortical nuclei by nearly ten to one. Thalamo-cortical projections are comparatively sparse, but serve at least two essential functions: (i) transmitting sensory inputs to the primary cortical areas (S1, A1, and V1), and (ii) providing a means to selectively amplify/synchronise cortex-wide activation.

GW theory also entails that conscious functions operate upon an information load about the size of working memory and therefore processing is at a highly coarse-grained level. In this context, global attention is (at least) a second-order operation, acting upon a highly selective stream of information. All this is to say that a relatively low density of widely distributed, yet highly convergent, circuits could be all which are required to create a conscious system; and these are the very characteristics of the neural model described.

However, most neural network modellers take a cortically-centred view of cognition, from which the brain stem functions so far described probably seem rather primitive or trivial (for example orienting, controlling eye movements) when compared to cortically-mediated processes such as language acquisition, pattern recognition, and motor planning. What evidence is there that cortical (and other forebrain systems) depend upon projections to the thalamus for effecting high-level cognitive processes?
Early support for such effects, mediated by prefrontal projections, was provided by animal experiments undertaken by Skinner and Yingling (1977). They found that selective activation of one portion of a fronto-thalamic tract could shut down sensory processing in visual, but not auditory, cortex. Activation of another ‘spoke’ of the prefrontal-thalamic tract shut down auditory processing, but allowed visual inputs to reach posterior cortex. This result implies that selective attention emerges via selective inhibition in certain sensory channels which the organism must know in advance are irrelevant to its situation. To inhibit orienting based upon advanced knowledge is clearly a sophisticated use of cognition.

In the brain, the primary areas (S1, A1, and V1) send no direct projections to PFC; but they do send convergent projections to secondary association areas, which send projections directly to PFC (as well as posterior association areas). Although these feed-forward projections to PFC are less topographically precise (for example the receptive fields of visual neurons in the secondary areas are much larger), they maintain a fair degree of parallel distribution, indicating that much of the prefrontal cortex is as modular in its organisation as the posterior ‘association’ cortex. Moreover, PFC ‘modules’ reciprocate these parallel, feed-forward projections, although in a more divergent pattern.

In PFC there are hundreds (if not thousands) of ‘central modules’. Feed-forward inputs allow them to use and store highly processed information from the posterior (sensory) cortex. Of course, feedback (or re-entrant) connections enable PFC to influence processing in the posterior areas as well. However, such divergent and indirect feedback pathways are poorly suited to exercising momentary, direct effects upon processing at the input level. Nor could such centrally-stored knowledge be directly employed to guide, or anticipate, how inputs are processed (‘knowing in advance’ Skinner and Yingling (1977) attributed to PFC-Th circuits). This is where direct projections to the primary processing areas (actually the thalamo-cortical circuit) prove quite valuable.

It is now generally accepted that the prefrontal lobes (with the cingulated cortex) constitute an ‘executive’ over the limbic system mediating such functions as working memory, inhibition of conditioned responses and goal-directed attention (Newman, 1997). More recent research on the basal ganglia have suggested that they constitute a ‘motor programming extension’ of the frontal lobes as well – routed through the thalamus. The BG ‘extension’ (like the thalamo-cortical loops)
sends rich, collateral projections to nRt which effect not only its ‘gating’ of motor programmes, but hippocampal-mediated episodic memory functions.

PFC acts as an executive attentional system by actively influencing information processing in the posterior cortex through its effects upon the nRt. In this manner, the highly parallel (processing) functions of the posterior cortex are brought into accord with increasingly complex and intentional cognitive schemes generated within the prefrontal regions of the brain. A defining property of an executive system is that it acts upon other sub-systems, modifying their inputs for its particular purposes. Posterior cortical areas act more like arrays of quasi-autonomous processing modules (or local experts). Note that an executive system is not an essential requirement for consciousness. That this is the case is illustrated by the literature on extensive damage to the frontal lobes of the brain. PFC damage results in significant deficits in such purposeful activities as inhibition of inappropriate responding; switching of response set, and planning and monitoring of actions, but produces little or no alteration in basic mental status. Indeed, many patients with frontal lobe pathology perform at pre-morbid levels on intelligence tests. In terms of the GW model presented, it is not executive attentional processes, but the selective binding of coalitions of active cortical modules via a thalamo-cortical competition which is the backbone for the generation of a coherent stream of conscious representations.

Finally, cortico-thalamic projections to nRt and associated specific nuclei are both more topographically precise and more pervasive than had once been thought. Llinas and Pare (1991) estimate that, for every axon the thalamus sends to the cortex, the cortical area projects to reciprocates with ten. Given the modular architecture of the neocortex, one might reasonably predict that these cortico-thalamic projections exert highly differentiated influences upon the flow of information through the thalamus. For example, recent and perhaps consistent findings indicate that the visual cortex appears to have a major action down onto the lateral geniculate nucleus, which may generate thalamic oscillations.

While additional research is clearly needed, current findings suggest that Scheibel’s (1980) early model of the converging influences of projections upon a thalamic hub – with the addition of basal ganglia inputs to nRt and the intralaminar complex – remains a viable model for ‘global attention’, including the influences of cortically generated ‘schemas, purposes and plans’. Newman (1997) discusses the contributions of the ‘cortico-basal ganglia-thalamo-cortical loop’ to memory and volitional processes in greater detail, while Taylor and Michalis (1995), among others, have developed neural models simulating functions of the BG and hippocampal systems.
Conclusions

A collection of neuroscience models for attention and binding, and resource allocation, which share important features and parallels with a Neural Global Workspace System for conscious attention (Newman and Baars, 1993), has been introduced. While the ‘wagon wheel’ model largely neglects the influences of memory and affective systems upon the stream of consciousness, the outlines of a general framework for understanding conscious processes should be discernable. This is certainly advancement, given the virtual terra incognita consciousness has been for most of the history of science.

B.5.3 Temporal binding and neural correlates of sensory awareness (Engel and Singer, 2001)

Introduction

Theories of binding have recently come into the focus of the consciousness debate, particularly the potential relevance of temporal binding mechanisms for sensory awareness. Neural synchrony with a precision in the millisecond range may be crucial for conscious processing, and may be involved in arousal, perceptual integration, attentional selection and working memory. Recent evidence from both animal and human studies demonstrates that specific changes in neuronal synchrony occur during all of these processes and that they are distinguished by the emergence of fast oscillations with frequencies in the gamma-range.

A large body of neuropsychological and physiological evidence suggests that consciousness has to be understood as a function of numerous interacting systems, such as sensory areas, memory structures, centres for executive control as well as circuits mediating emotion and motivation (Delacour, 1977; Young and Pigott, 1999). Thus, any theory about the neural correlates of consciousness (NCC) must explain how multiple component processes can be integrated and how large-scale coherence can emerge within distributed neural activity patterns. Furthermore, such a theory must specify mechanisms for the dynamic selection of subsets of neuronal responses, because only a fraction of all available information gains access to consciousness. It is suggested that achieving both cross-systems coherence and dynamic response selection requires mechanisms for binding of distributed information (Engel et al., 1992; Singer et al., 1997).
There seems to be wide agreement (Crick and Koch, 1990; Newman and Baars, 1993; Tononi and Edelman, 1998) that the physiological prerequisites of awareness include: (i) arousal: the 'waking up' of the brain by non-specific modulatory systems, (ii) sensory segmentation: the basic step in sensory processing, which comprises both detection and binding of object features, (iii) selection: that is, processes (including attention) which lead to an enhanced efficacy of subsets of neural signals, and (iv) working memory: the short-term storage of information about the current situation. It is proposed that all these processes either require or modify the operation of neuronal binding mechanisms. Moreover, the binding processes relevant for the instigation of awareness may be implemented in the temporal domain, that is, by transient and precise synchronisation of neuronal discharges (Engel et al., 1999).

Binding and consciousness

The notion of binding has been introduced first in the context of feature integration (Treisman, 1996) and perceptual segmentation (Von der Malsburg, 1981). Subsequently, the concept of binding has been applied to other domains and is now employed in theories on object recognition (Hummel and Biederman, 1992), attention (Niebur et al., 1993), memory formation and recall (Damasio, 1990), motor control (Murthy and Fetz, 1992), sensorimotor integration (Roelfsema et al., 1996), language processing (Pulvermüller, 1999) and logical inference (Shastri and Ajjanagadde, 1993). In all these domains, a set of related computational requirements has been identified which, taken together, define what has been termed the ‘binding problem’ (Treisman, 1996):

a) Information processing underlying the functions listed above is distributed across many neurons spread out over different areas or subsystems and, thus, neurons currently participating in the same cognitive process need to be ‘tagged’. This, in turn, requires a mechanism for the expression of specific relationships between individual neural signals.

b) Perception of and action in a complex environment usually requires the parallel processing of information related to different objects or events which have to be kept apart to allow sensory segmentation and goal-directed behaviour. Thus, neuronal activity pertaining, for example, to a particular object needs to be distinguished from unrelated information in order to avoid confusion and erroneous conjunctions (Von der Malsburg, 1981).

c) It has been claimed that specific yet flexible binding is required within distributed activation patterns to allow the generation of syntactic structures and to account for the systematicity and productivity of cognitive processes (Fodor and Pylyshyn, 1988).
Most cognitive functions imply the context-dependent selection of relevant information from a richer set of available data. It has been suggested that appropriate binding may be a prerequisite for enhancing the saliency of subsets of responses and, thus, for further joint processing of those signals pertaining to some particular contents (Singer et al., 1997).

The central hypothesis is that these facets of the binding problem also apply to the issue of consciousness (Metzinger, 1995) and that, hence, unravelling the mechanisms capable of solving the binding problem may be critical for understanding the NCC. Damasio (1990) has suggested that conscious recall of sensory contents requires the binding of distributed information stored in spatially separate cortical areas. Llinás and Ribary (1990) have argued that arousal and awareness result from the activation of non-specific thalamo-cortical circuits which serve to bind sensory contents encoded by specific thalamo-cortical loops. Similarly, Newman and Baars (1993) have suggested that unspecific and specific thalamo-cortical systems interact to form a 'global workspace', where bound contents become globally available and, hence, lead to the emergence of conscious states. Finally, Grossberg (1999) has proposed that conscious states result from a resonance, or match, between top-down priming and bottom-up processing of incoming information, which allows learning and binding of information into coherent internal representations.

Taken together, all these authors seem to imply a set of common assumptions, namely: (i) that consciousness results from a cooperative process in a highly distributed network, and is not attributable to a single brain structure or process, (ii) that binding is highly relevant for the NCC, and (iii) that only coherent activity, resulting from the operation of binding mechanisms, could become functionally salient, causally efficacious and globally available, and, thus could lead to the emergence of conscious mental states and their respective behavioural manifestations. The critical point is that binding may not only serve for achieving the 'unity of consciousness' but, first of all, for 'gating' the access to awareness and, hence, for turning subconscious information into conscious mental content.

Temporal binding

The concept of dynamic binding by synchronisation of neuronal discharges has been developed mainly in the context of perceptual processing. The most dramatic case is represented by the primate visual system where anatomical and physiological studies have led to the identification of more than 30 distinct visual cortical areas (Felleman and Van Essen, 1991). This parcellation is
assumed to reflect some kind of functional specialisation because neurons in each of these visual areas are, at least to some degree, selective for characteristic subsets of object features. As a consequence of this functional specialisation, any object present in the field of view is represented by a large distributed set of activate neurons – a so-called cell assembly.

On theoretical grounds, it has been suggested that the binding problem arising in distributed networks may be solved by a mechanism which exploits the temporal aspects of neuronal activity (Von der Malsburg, 1981). The prediction is that neurons which represent the same object or event might fire their action potentials in temporal synchrony with a precision in the millisecond range. Synchrony would selectively tag the responses of neurons which code for one object and demarcate their responses from those of neurons activated by other objects (Von der Malsburg, 1981). This highly selective temporal structure would allow the co-activation of multiple assemblies in the same network which nonetheless remain distinguishable. Moreover, temporal binding could serve as a mechanism for selection of assemblies for further processing, because precisely synchronised spikes constitute highly salient events which can be detected by coincidence-sensitive neurons in other brain areas (Abeles, 1982; König et al., 1996).

Crick and Koch (1990) applied this concept of temporal binding to the issue of consciousness. Inspired by the finding that visual stimuli can elicit synchronised activity in the visual cortex, they proposed that an attentional mechanism could induce synchronous discharges in selected neuronal populations, and that this temporal structure would facilitate transfer of the encoded information to working memory. At the time it was published, Crick and Koch’s speculative proposal was not supported by experimental evidence, however, more recent results suggest that temporal binding may indeed be a prerequisite for the emergence of awareness.

Animal studies on synchrony and awareness

By now, the synchronisation phenomena predicted by the temporal binding hypothesis have been documented for a wide variety of neural systems. Characteristic features of the synchronisation observed can be summarised as follows:

a) In all systems and species studied, synchrony can be very precise, with a coincidence window of about 10 msec;

b) Synchrony reflects the topology of feature space and is clearly dependent on, for example, proximity of receptive fields and similarity of neuronal feature preferences.
c) Synchrony can occur both internally generated (non-stimulus-locked) as well as externally imposed (stimulus-locked). The former type occurs predominantly with responses to stimuli lacking a distinct temporal structure or with self-generated activity and is due to interactions mediated by reciprocal intrinsic connections (Engel et al., 1991). The latter, in contrast, is characterised by phase-locking to the stimulus, occurs in response to rapid stimulus transients (Rager and Singer, 1998) and is due to synchronised sensory input signals.

d) In many studies, the synchrony observed was associated with an oscillatory modulation of the responses at frequencies in the gamma range, that is above 20 Hz. These oscillations can occur both with internally generated synchrony and in stimulus-locked activity (Steriade, 1968).

Functional relevance of synchrony

As the animal studies indicate, synchrony relates to all four presumed component processes of awareness, namely, arousal, segmentation, selection and working memory. Clearly, precise synchronisation of neuronal discharges is more prevalent during states characterised by arousal, and moreover, gamma-oscillations are also particularly prominent during epochs of higher vigilance (Steriade et al., 1996; Munk et al., 1996). Furthermore, several lines of evidence make it likely that temporal binding is highly relevant for scene segmentation, leading to structured representation of sensory inputs. A key observation supporting this notion is that neuronal synchronisation depends on the stimulus configuration. Thus, spatially separate cells show strong synchronisation only if they respond to the same object. However, if responding to two independent stimuli moving in different directions, the cells fire in a less correlated manner or even without any fixed temporal relationship. This effect has been documented for cortical (Gray et al., 1989; Freiwald et al., 1995; Brosch et al., 1997; Castelo-Branco et al., 2000) and subcortical (Brecht et al., 1999) neurons in anaesthetised cats as well as for cortical cells in anaesthetised (Livingstone, 1996) and awake, trained (Kreiter and Singer, 1996) monkeys.

The idea that binding of neural activity by synchronisation might activate working, or short-term, memory has been a key ingredient of Crick and Koch’s (1990) awareness model. They speculated that synchronised activity might either reverberate in neural circuits, or trigger short-term changes in synaptic efficacy to express transient memory states in the cortex. As indicated by both experimental studies and computational modelling approaches, volleys of synchronised discharges may be particularly efficient in creating reverberatory activity in cell assemblies, which might be a
basis for transient memory (Abeles et al., 1993; Diesmann et al., 1999). Several recent studies also suggest a role of synchronised discharges for synaptic modification. In-vitro studies have demonstrated that temporal relations between pre- and postsynaptic activity are critical for the occurrence of long-term potentiation (LTP) or long-term depression (LTD) with a coincidence window of 10-20 msec which is defined by temporal continuity with the spike backpropagating into the dendrites of the postsynaptic cell (Markram et al., 1997). Oscillations may be of particular relevance for memory formation, because they establish ‘windows of depolarisation’ which may be critical for LTP and LTD to occur (Huerta and Lisman, 1995).

Relationship to perceptual selection

Recent evidence also indicates that synchrony may be relevant for the selection of sensory information for access to awareness. This is suggested by a study of Fries et al. (1997) in which neuronal responses were recorded from cat visual cortex under conditions of binocular rivalry. This study investigated the hypothesis that response selection in primary and secondary visual cortex might be achieved by modulation of the synchrony rather than the rate of discharges. The results showed that neurons representing the perceived stimulus were strongly synchronised, whereas cells processing the suppressed visual pattern showed only weak, if any, temporal correlation. In these experiments, synchrony across recording sites was accompanied by prominent gamma-oscillations, which showed the same changes under the rivalry condition: the power in the gamma-band increased for neurons representing the dominant stimulus, although it decreased for cells responding to the suppressed pattern. Importantly, however, no differences were noted under the rivalry condition for the discharge rates of cells responding to the selected and the suppressed eye, respectively (Fries et al., 1997). These results therefore demonstrate that, at least at early processing levels, dynamic selection and suppression of sensory signals are associated with modifications of the synchrony rather than the rate of neuronal discharges. Changes in temporal correlation patterns at early stages of processing should result in changes of discharge rate at later stages, if the saliency of responses depends on their synchronicity. Studies in awake monkeys are in accordance with this prediction, showing that unequivocal rate changes under perceptual rivalry are observed predominantly in higher cortical areas (Leopold and Logothetis, 1999).

Additional support for a relationship between temporal binding and perceptual selection comes from recent work on the superior colliculus, a midbrain structure with important integrative functions which mediates orienting responses towards a target of interest. Findings suggest that potential targets for orienting behaviour are represented in the colliculus by assemblies of
synchronously firing cells. More recent experiments have attempted to test directly the idea that temporal binding plays a role in target selection in the colliculus (Brecht et al., 1997). These data strongly suggest that synchrony in the millisecond range is an important determinant for target selection in the cortico-tectal pathway.

Taken together, the results from animal studies suggest that synchronisation is involved in the generation and maintenance of awareness. These experiments show that synchrony and gamma-oscillations occur in a state-dependent manner, are related to stimulus coherence, co-vary with perceptual selection, and have relevance for the formation of memory traces. This in turn suggests that temporal binding mechanisms may be important for the processes of arousal, segmentation, selection and working memory, respectively. Clearly, these results do not imply that synchrony would be the only mechanism for binding and selection relevant for the NCC. Changes in firing rates, as the other highly relevant coding dimension, can also contribute to an enhancement of the saliency of neural signals (Shadlen and Movshon, 1999).

**Human studies on timing and perceptual consciousness**

In humans, numerous studies using the technique of electroencephalographic (EEG) or magnetoencephalographic (MEG) recording have provided evidence supporting the conclusions drawn previously. An important methodological difference is that the signals recorded in EEG/MEG studies result from spatial averaging across large neuronal assemblies and, thus, at this macroscopic level the cellular processes of synchronisation and oscillatory response structure cannot be dissociated. Rather, both phenomena show up in a lumped fashion as changes of power in particular frequency bands.

As other mammals, humans show an enhancement of high-frequency EEG components during states of increased arousal, sleep-waking transitions and REM sleep. Several studies indicate that in the awake state and during REM sleep, gamma-band frequencies are present in the EEG or MEG, which are diminished during deep sleep (Llinás and Ribary, 1993). The similarity of high-frequency activity during REM phases and the awake state has led to the suggestion that, in both cases, synchrony in this frequency band correlates with similar processes leading to consciousness, which are just differently modulated by external stimulation (Llinás and Ribary, 1994). During wakefulness, synchronisation in the gamma-band is enhanced in epochs of phasic arousal which may occur in response to alerting stimuli and during orienting or investigatory action (Sheer, 1989). Tallon-Baudry et al. (1996) were among the first to show in humans that perception of coherent
objects is specifically associated with ‘induced’ gamma activity which is not phase-locked to stimulus transients.

The relation of synchrony to perceptual selection has been investigated in several recent studies. Electrical (Brown and Norcia, 1997) and neuromagnetic (Srinivasan et al., 1999) correlates of binocular rivalry were investigated using the method of ‘frequency tagging’, that is, the stimuli presented to the two eyes were flickered at different frequencies. Under these conditions, cortical steady-state responses are dominated by the two stimulation frequencies, and it was shown that the power in the frequency band designating the left- or right-eye driven assembly increases in epochs where the respective eye has contributed to perception (Brown and Norcia, 1997; Srinivasan et al., 1999). Moreover, perception of a stimulus is associated with an increase in intra- and interhemispheric coherence of neuromagnetic signals at the stimulus frequency (Srinivasan et al., 1999). These data show that synchrony in neuronal assemblies can be modified as a function of perceptual state. In agreement with the study on binocular rivalry in cats discussed in the preceding section, this suggests that temporal binding mechanisms may contribute to selection of signals for access to awareness.

Finally, evidence from studies in humans suggests that working memory, as another prerequisite for the emergence of consciousness, may also depend on temporal coordination of neuronal populations. The relationship between gamma-band synchrony and working memory has been investigated in a study by Tallon-Baudry et al. (1998), who have shown that during a visual delayed-match-to-sample task changes occur specifically in the frequency band between 25-60 Hz, indicating an enhancement of precise synchronisation over ventral occipital and prefrontal areas. A study by Sarnthein et al. (1998) has also reported increased coherence between prefrontal and posterior electrodes during a visuospatial working memory task, which was observed in the gamma-band but interestingly also at lower (theta) frequencies.

Conclusions: synchrony and conscious states

The studies reviewed above strongly suggest that the temporal dynamics in neuronal activity may be critical for the production of conscious states. The experiments on binocular rivalry make it very likely that only strongly synchronised neuronal signals contribute to awareness. They suggest that activation of feature-detecting cells is per se not sufficient to grant access of the encoded information to consciousness (as indicated by the fact that cells representing a non-perceived stimulus are still well responding – Fries et al., 1997). The processes of arousal, segmentation,
selection and working memory may together form (part of) the neural correlate of awareness. All four processes relate to the operation of temporal binding mechanisms which, thus, may constitute a critical component of the NCC.

As mentioned above, arousal is characterised by an enhanced precision of neuronal synchrony and a shift to high oscillation frequencies, indicating that thalamo-cortical systems change from large-scale synchrony into states with more specific, regionalised temporal patterning. Central activating systems may act to modify, in a task- and context-dependent manner, the efficacy of temporal binding mechanisms. This may change both the spatial range and the specificity of neuronal interactions and, thus, contribute to more specific information processing.

As an additional prerequisite, consciousness requires the completion of basic sensory processing steps, including feature detection and segmentation. The latter seems necessary because it is impossible to extract meaning out of sensory information without prior structuring. Although encoding of object features is presumably achieved by rate modulation in single neurons or across populations, segmentation requires dynamic binding. Segmentation may be implemented in the temporal domain. Synchrony as a binding mechanism allows establishing specific relationships between neural discharges which are, in principle, independent of spatial proximity or direct neuronal connections. Synchrony is not only determined by the stimulus, but is modulated in a context- and task-dependent way by cooperative interactions within the cortical network. This leads to functional coherence among neurons which convey relevant information.

Current awareness theories assume that not all of the computational results of sensory processing contribute to consciousness. Rather, as an additional step, part of the information is subjected to a selection process which 'gates' access to awareness. Selection can be mediated by neural synchronisation, as temporal coincidences are more easily 'detected' by other neural assemblies than temporally dispersed signals. Only activity patterns carrying a strong temporal signature may be functionally efficacious and globally available and, therefore, such a signature may be a fundamental prerequisite for making information available to other brain centres. The selection is controlled both by bottom-up (for example stimulus novelty) and top-down (for example expectancy, memory contents) influences, which can lead to competition among different assemblies and result in changes of synchrony.

It should be emphasised at this point that the notion of selection, as employed here, is broader than the notion of attention. Clearly, there are cases of response selection (for example procedural
selection occurring during sensorimotor coordination) where very specific binding can occur without any awareness. The notion of attention, by contrast, always implies access to consciousness, and it may be viewed as the extreme case of very specific, ‘very-serial’ selection of episodic contents. As such, attention is not a ‘force’ which induces, or creates, synchrony (Niebur et al., 1993) but a process which itself may be implemented by dynamic changes in temporal relationships.

Finally, synchrony may be ideally suited to promote access of selected contents to working memory. Synchronised assemblies may transiently stabilise in some reverberatory state, endowing them with competitive advantage over temporally disorganised activity. This may provide the basis for working memory necessary to achieve the ‘holding’ of situational context in the respective processing areas. The information carried by such assemblies during working memory states may become conscious.

Moreover, the model advocated here may account for the ‘global availability’ (Von der Malsburg, 1997) of conscious information, because temporal signatures which reliably propagate across systems may be suited to achieve coherence, or resonance, between assemblies in different neural systems (as required, for instance, during action planning (Roelfsema et al., 1996) or language processing (Pulvermüller, 1999)).

Beyond sensory awareness, the temporal binding model could have implications for higher-order consciousness processes which, in addition, seem to require the activation of motivation and action planning systems, of episodic memory and, eventually, symbol processing capacities. In all likelihood, these faculties will require cross-modal and cross-system binding. Temporal binding may establish patterns of large-scale coherence, thus enabling specific cross-system relationships which bind subsets of signals in different modalities.

An important point to be mentioned is that large-scale coherence is not equivalent to uniform synchrony. Indeed, global synchronisation is associated with a low complexity of neural interactions which, as seen in deep sleep or epilepsy, is counterproductive to consciousness (Tononi and Edelman, 1998). Consciousness may require the embedding of contents into progressively higher-order contexts, both in space and time. This recursive embedding might be mediated by hierarchical binding of assemblies into higher-order arrangements which could be achieved, for example, by multiplexing of interactions in different frequency bands. Such higher-order bindings
could form the basis for 'meta-representations' necessary to incorporate low-level contents into global world- and self-models.

In the current debate (Reynolds and Desimone, 1999; Ghose and Maunsell, 1999; Treisman, 1999) about the significance of binding and its possible mechanisms, objections can, in principle, rely on two arguments: first, it could be denied that binding problems to exist altogether; and second, acknowledging the binding problem, one could still reject the idea that temporal binding may provide a theoretically viable or physiologically plausible solution (Shadlen and Movshon, 1999). With respect to the NCC, the first option seems a hard choice, because consciousness constitutes, without doubt, an integrative process par excellence. Choosing the second option presupposes the contemplation of alternative binding mechanisms. However, as discussed elsewhere in depth (Singer et al., 1997), it is not easy to find adequate candidates. In the case of consciousness, anatomical convergence onto 'higher-order' neurons or grouping of signals by place codes may be too inflexible, and attention (Treisman, 1999) itself may be part of the problem, rather than part of the answer (Engel et al., 1999).

Some questions remain. For example, why does selection based on temporal binding lead to awareness in some cases but not in others? For instance, both the attentive search for a particular object as well as the visuomotor coordination in a frequently practised task such as driving require context-dependent selection. However, although in the former case the selection process usually leads to awareness, this does not necessarily hold for the latter. What makes the difference?
APPENDIX C

CONNECTIONIST MODELLING OF COGNITIVE PROCESSES

C.1 Overview of Connectionism

C.1.1 Challenges of connectionist modelling

The brain of a newborn child contains billions of neurons, but the child can perform virtually no cognitive functions. After a few years, receiving continuous streams of signals from the outside world via his/her sensory surfaces, the child can see, understand language and control the movements of his/her body.

As an adult, a person can see, walk and understand speech without conscious effort, but has no memory of a time when he/she could not. Nor can the person remember the process by which these tasks were learnt. It is difficult to imagine that there is any problem in learning to see or to understand speech. To understand how extraordinary this achievement is, consider what the infant needs to do to recognise words in continuous speech. If spoken words were presented in isolation and each example of a word was the same, as it is in print, it might not be too difficult for a child to discover that these units were significant. However, this is far from the case as words in speech do not come with gaps around them. One hears isolated words by recognising the words which the speaker used. In the original signal, the sound patterns for most words run into each other without breaks. The separation of the speech stream into successive words is a construction of the speech recognition system.

If words were always the same it might seem possible for the child to recognise these regular units even if they came in a continuous stream. However, the signals arriving at the ear corresponding to a given word are quite different when produced by different speakers, or by the same speaker in a different context. Even when a word comes from the same speaker in the same context it will seldom be the same signal at the ear because speech is superimposed on all the other random background noises in the environment. If the child had been given the concept of ‘word’, and a specification of its duration and properties, one could imagine the child starting to extract some
regularities from the speech stream. However, nobody would start to teach a child to understand speech by telling the child what a word was. Without any idea of what to look for, or even that there is something worth looking for, how could the regularities which are hidden in speech be discovered? There is such infinite variety in the input that it seems impossible for the infant to get started. Why should the brain ever discover that the word was a unit when the pattern of nerve firing arriving from the ear could be grouped in so many ways? Yet, every child does discover how to extract words from the endless variety of speech in the first few years of life.

Such problems in understanding speech, and ones of similar difficulty in learning to see or to control movement, have, for years, seemed quite intractable. However, in the last decade there has been a revolution in the understanding of how a system like the brain, given streams of input from a structured world by its sensory receptors, could come to discover that structure. Part of this revolution has been the discovery of how cognitive functions could be performed by a system which computes with simple neuron-like elements, acting in parallel, on distributed representations. The investigation of what can be achieved by models which perform parallel distributed processing is called connectionism.

In this appendix, a review by McLeod et al. (1998) of connectionism is summarised.

C.1.2 Capabilities of connectionist modelling

Connectionist models have given a precise match to data obtained in experiments with human subjects. For example, in experiments which measure the speed at which adult readers can read individual words aloud, it is found that the time taken depends on both the frequency of the word and the regularity of its pronunciation pattern. Connectionist simulations do not just mimic results which are already known. Their predictions have suggested fruitful areas for experimentalists to investigate.

For example, when children start producing the past tense of verbs they initially get them correct and then they begin to make errors. A child who correctly uses the irregular form ‘went’ to express the past tense of ‘go’ might say ‘goed’ a few months later. Eventually the child will revert to using ‘went’ correctly. (Adding –ed to the verb stem is the way that regular past tenses are formed in English, so ‘goed’ is called an over-regularisation error.) This paradoxical behaviour has been cited as evidence for the theory that linguistic development proceeds in stages. In this case, the argument goes, the stages are: (i) the child learns individual words (such as ‘went’), (ii) then
acquires rules such as ‘add -ed to form the past tense’, and (iii) discovers that rules have exceptions.
A connectionist simulation of children acquiring the past tense by Plunkett and Marchman (1996) 
produced over-regularisation errors, like children, but it did not produce stage-like behaviour.
Irregular verbs were not all regularised once over-regularisation appeared, as the stage theory would 
predict. The discrepancy between the model predictions of the model and the standard theory sent 
experimentalists out to discover the fine detail of the pattern of production of over-regularisation 
errors in children.

Connectionist simulations have suggested solutions to some of the oldest problems in cognitive 
science. For example, if one knows a particular person, it is possible to recognise the person from 
whatever angle one sees him/her. This is such a commonplace everyday experience that it may be 
hard to realise that the achievement is in any way remarkable. In fact, how to form a representation 
of an object which will permit recognition from any viewing position, is a notoriously difficult 
problem. The image produced by someone seen in profile has little in common with that produced 
when seen full face. How can one form a representation of somebody else which will allow one to 
recognise the person in future when he/she presents a different image? Connectionist models have 
shown ways that this can be done. The test for view invariance is whether units in the output layer 
of the model respond to one particular face independent of the view which it is shown. If so, then 
the net has built up representations which have discovered what is in common between different 
images of the same person even though the images themselves have little in common.
C.2 Principles of Connectionism

C.2.1 Basics of information processing

A casual glance at connectionist research literature may give an impression of dauntingly complex models, requiring advanced mathematical knowledge to be understood. Despite the complexity of the behaviour they can simulate, the principles on which connectionist models are based are simple. These are derived from observations of the organisation of information processing in the brain:

a) The basic computational operation in the brain involves one neuron passing information related to the sum of the signals reaching it to other neurons.
b) Learning changes the strength of the connections between neurons and thus the influence that one has on another.
c) Cognitive processes involve the basic computation being performed in parallel by large numbers of neurons.
d) Information, whether about an incoming signal or representing the memory of the network of past events, is distributed across many neurons and many connections.

The central principles of connectionist models are derived from current knowledge of computation within the brain, so the models are said to be ‘neurally inspired’. This puts them in stark contrast to traditional models in cognitive psychology or Artificial Intelligence (AI). Traditional models in cognitive psychology contain elements like limited capacity channels, articulatory loops and short-term memory stores. Models in AI contain sets of rules. Either approach allows modelling of human cognitive capacities, but in general no attempt is made to relate the operations these elements perform to the way the brain works at a neuronal level. The aim of a traditional model is to describe the performance of the subject, not the way that the performance is achieved. The connectionist approach is the reverse. It starts with a model that incorporates brain-like processing and sees whether behaviour emerges which mimics that shown by people.

Since the brain does compute with neurons it might seem to go without saying that neural plausibility is a virtue in modelling cognitive processes. Nevertheless, there are some drawbacks to this approach. Researchers do not have a complete understanding of the way that information is transmitted between neurons in the brain. The method of passing information from neuron to neuron in connectionist models certainly occurs, at least at a conceptual level. However, in the brain there are other methods of inter-neuronal signalling which are not yet implemented in connectionist models. In addition, some connectionist models contain procedures, especially
learning algorithms, which are at present believed not to occur in the brain. Completely accurate neural models may be possible in the future. At the moment neural plausibility is a relative rather than an absolute virtue. Contemporary connectionist models are based on the assumption that although they make simplifications, they can provide a useful starting point for understanding how cognitive computations might be performed.

An appealing aspect of the connectionist approach is that the models of cognitive processes are computational. That is, they actually produce a response to a stimulus. The predictions that such models make about reaction time or error rate or interference can be compared at a quantitative level to the behaviour produced by subjects in experiments. The inability of many traditional models of cognitive processes to do more than make qualitative predictions about the effect of some experimental variable on performance has often made it difficult to choose between them because they make the same qualitative prediction.

Assumptions about computation in the brain on which connectionist models are based

Although much of the detail about computation by the brain is not yet understood, these principles are at a sufficiently abstract level that they seem a plausible starting point:

a) Neurons integrate information. Many different types of neuron have been discovered. Three are shown in Figure C.1. Despite their bewildering variety in detail, they perform a common function – they integrate information about the firing of one set of neurons (their input) and pass information related to this input (their output) to a new set of neurons. In a ‘classical neuron’, a simplified approximation which captures the essence of the function of a real neuron, this operation takes place in three stages. First the neuron receives signals, either excitatory or inhibitory, from other neurons via synaptic connections onto its dendrites. If the sum of these signals exceeds a threshold, the neuron fires. This is communicated to other neurons by a signal passing down its axon. This signal acts in turn as part of the input to the dendrites of other neurons. Connectionist models consist of a number of these simple units. Each line coming into the unit from above represents an input connection. These connections may be positive or negative, so activity on an input line may increase or decrease the activity of the central unit. The central unit sums the inputs and passes information about the sum down the output connections to the other units to which it is connected. Thus, the functional role of a unit in a connectionist model is the
same as that of a classical neuron – it passes information about the pattern of activity of one set of units to another set.

**Figure C.1** Three types of neuron: from left to right, a spinal motoneuron, a hippocampal pyramidal cell, and a Purkinje cell of the cerebellum; the last figure is a computational unit in a connectionist model (McLeod et al., 1998)

b) **Neurons pass information about the level of their input.** The output of a neuron communicates more than just the fact that it is receiving input. Its output varies systematically to convey information about the level of its input. In the classical neuron, information about the level of stimulation of the sending neuron is coded by the rate at which it fires. Information transmission between connectionist units achieves the same end, conveying information about the level of input, but in a different way. Each unit has an activity level. The activity level is transmitted as a single value – not as a rate of production of pulses – to all the units to which it is connected. The activity level is related to the input level – the higher the input, the higher the activity.

c) **Brain structure is layered.** Information is processed in the brain by a flow of activity passing through a sequence of physically independent structures. The organisation of part of the visual processing system is shown in the upper part of Figure C.2. If a word is presented to a child who is learning to read aloud, a pattern of excitation passes through a series of different structures. First the word stimulates receptors in the retina at the back of the eye. This produces a pattern of activity which passes along the optic nerve to the lateral geniculate body (LGN – part of the thalamus), a collection of cells about the size of a peanut in the middle of the brain. The pattern of firing from these passes along a bundle of fibres called the optic radiation to the visual cortex, a layer of cells a few millimetres thick at the back of the brain. The output from the visual cortex passes to a variety of other parts of the
cortex, such as the motor cortex, which transform it in various ways, until it finally becomes a signal which stimulates the vocal muscles to produce an appropriate sound. At the retina a pattern of activity is produced by an external stimulus. At each subsequent stage, namely the lateral geniculate body, visual cortex, motor cortex, there are millions of interconnected neurons which perform a transformation of the information carried by the incoming stream of nerve pulses.

![Diagram of the layered structure of information processing in the brain and connectionist models](Image)

**Figure C.2** The layered structure of information processing in the brain and connectionist models (McLeod et al., 1998)

The first few layers of this transformation from a pattern on the retina to activation of vocal muscles are represented in Figure C.2 by the conceptual brain underneath the real one. This view of the organisation of information processing in the brain is represented in connectionist models. The bottom part of Figure C.2 shows a typical multi-layered connectionist model which is simulating reading aloud. The row of units on the left is the input layer where the stimulus is presented to the network. The pattern of activity here will
vary with the letter string presented, just as the pattern of activity on the retina is different for different letters. This would enable the network to distinguish between different words just as the early stages in the visual pathway of a real brain would convert the pattern of stimulation on the retina into a pattern of firing which distinguishes one letter from another. Units in the input layer are connected to units in the middle layer. Units in the middle layer are in turn connected to units at the right, the output layer. The units in the output layer correspond to sound features. The units in the middle layer which neither receive input directly from the stimulus nor produce the output are called hidden units. Most models of psychological processes contain at least one such layer. They have a vital role to play in allowing networks to find solutions to difficult (non-linear) problems.

d) **The influence of one neuron on another depends on the strength of the connection between them.** A typical neuron in the brain influences the firing of several thousand other neurons. The influence takes place at a connection, called a synapse, between the axon of one neuron and the dendrite of another. When the axon on one side of the synapse fires it makes the neuron on the other side either more or less likely to fire. The effect of one neuron on another (whether it makes it much more or less likely to fire or whether it only slightly changes the probability) is determined by the strength of the synaptic connection between them. This is represented directly in connectionist models. The effect that one unit has on another is determined by the strength of the connection between them and is called the weight of the connection.

e) **Learning is achieved by changing the strengths of connections between neurons.** Experience can change the behaviour of an organism in response to a particular stimulus. Although the mechanisms by which this happens in the brain are not fully understood, there is good evidence that learning involves changing the strength of synaptic connections between neurons. Learning is implemented in connectionist models by rules which determine how the weights of the connections between units are changed. This idea is central to connectionism’s contribution to understanding cognitive processes. A fundamental aspect of many connectionist models is that they are models of learning. They try to explain how a particular set of experiences could lead an initially unstructured system to the acquisition of knowledge.

A model, such as the one shown in Figure C.2, would start with a random set of weights (or strengths) for the connections between units in the different layers. When a stimulus is first
presented to the network, the pattern of activity on the output units would be random, just like a child’s. Just as with the child, an external teacher then tells the network what the correct behaviour is. The network learns from this experience by changing the weights of the connections between the input and output units in such a way that next time the same stimulus is presented as input it will produce an output which is closer to the correct behavioural pattern.

Intuitively it might seem implausible that such simple structures could ever do anything interesting, let alone simulate the richness of human intelligence. However, as Crick (1989) mentioned, the results which can be achieved with neural networks are astonishing.

C.2.2 Attraction of parallel distributed processing for modelling cognition

The last decade of the 20th century has seen an explosive growth in the connectionist modelling of cognitive processes, with simulation of most of the classical experimental paradigms of cognitive psychology. One reason for this enthusiasm is that, independent of their success at modelling human performance at any particular cognitive task, all connectionist models exhibit some general characteristics which are shown by human cognitive processes and distinguish them from non-biological computational systems such as computer programmes – they still perform reasonably well after minor damage to components of the system, they still perform reasonably well if their input is noisy or inaccurate, and they allow memory retrieval by content.

Like the principles of inter-neuronal communication described, these characteristics of connectionist systems are based on general observations of brain structure. First, knowledge representation is distributed across many processing units. Second, computations take place in parallel across these distributed representations. The result is that conclusions are reached on the basis of a consensus of many calculations rather than depending on any particular one.

These principles put connectionist models in direct contrast to many traditional models in cognitive psychology or AI where knowledge representation is local and computation is serial. In general, such models are not immune to damage or resistant to noisy input. So a traditional model of, say, syllogistic reasoning, might give as good a fit to the experimental data as a connectionist model, but it would do so without exhibiting the full range of human characteristics as it performed the task.
Distributed representation of knowledge in connectionist network

The information storage systems well-known today, for example dictionaries, telephone directories, and computer discs use local representation. Each discrete piece of information is stored separately.

In connectionist models information storage is not local, but distributed. There is no one place where a particular piece of knowledge can be located. All the knowledge that the network contains is superimposed on the same set of connections. Although intuitively this may seem entirely implausible, the same set of weights store independent and even contradictory pieces of information. Some of the emergent properties of such systems are intriguingly similar to properties of human cognitive processes.

Distributed representations are damage resistant and fault tolerant

When one considers the structure of the brain, it is remarkable that it ever manages to come to the correct conclusion about anything. By any conventional standards neurons are an entirely unsuitable medium for computations – they die throughout the brain’s life, causing random loss of stored information, they have a finite probability of firing even when they are not engaged in signal processing, and the response of a neuron to any particular input is probabilistic, not fixed.

If the processing components in a conventional digital computer produced random spontaneous output, a different response to the same stimulus on different occasions would be possible and the computer could suffer from random component drop-out, causing the system to be totally unpredictable. Sometimes it would work correctly, but if a computation required access to the contents of a missing memory unit or a burst of noise obliterated a signal, the result would be garbage. Although the components in the brain can fire and die at random, the computations performed by the brain are not unpredictable. With minor damage, it becomes a little slower and less accurate, but it still produces roughly the same answer. It has to suffer major damage before it produces nonsense.

The brain escapes the consequences of the unpredictable behaviour of individual neurons because its computations are performed in parallel on representations which are distributed over many neurons. Even if individual components of the calculation are not accurate, the ensemble average can nevertheless give an answer which is accurate enough. When the memory location required for
a calculation in a localist information storage system is damaged, the result is disastrous. For example, if the first page of a dictionary is missing, there is no way of checking whether aardvark is really spelt like that. However, in a connectionist system, there is no such thing as 'the memory location required for a calculation'. Information and calculation are spread across the network. If one unit or connection in the network is damaged, others can make up for the missing part.

A neural system will be slightly less accurate if a connection is lost, but the pattern of loss is quite different in localist systems. Damage to a localist system causes some information to be lost totally while other information is unaffected. In a distributed system any damage causes partial loss of a range of information. As damage increases, the performance of the system inevitably begins to drop. However, a small amount of damage may have no noticeable effect on the output of the system. The ability of brains and connectionist models to continue to produce a reasonable approximation to the correct answer following damage, rather than undergoing catastrophic failure, is an example of fault tolerance referred to as graceful degradation.

Memory access by content in connectionist networks

Human memory is content-addressable allowing one to access a memory by using some part of the information contained in the memory (the content) as a retrieval cue. This is unlike retrieval from familiar forms of information storage, such as dictionaries, telephone directories or computer discs. In these, the place where the information is stored has an address.

One of the reasons why connectionist models of human memory are attractive is that content address-ability follows as a natural consequence of their distributed structure. Content address-ability can be built into localist storage but only by adding a complex cross-referencing system. An attractive aspect of content-addressable memory is that it is inherently fault tolerant. With content-addressable memory, the weight of evidence pointing to a particular answer can overcome other evidence that is inconsistent. A best-fit solution can be chosen even if it is not perfect. This is unlike memory systems in which access by address is the only possibility where any error in the address will lead to failure.

In a distributed connectionist system, an attempt to extract any information from the system leads to a flow of excitation and inhibition throughout the system to all connections which has a relation to this information. This results in many different nodes becoming active. What is retrieved is the information which corresponds to the most active mode(s) once this flow has stabilised. When
When activity flows through a connectionist network in response to an input, each unit influences the state of all the units to which it is connected. If the connection weight is positive the sending unit tries to put the receiving unit into the same state of activity as itself, and if the connection weight is negative it tries to put it into the opposite state. Since all activity changes are determined by these influences, each input can be seen as setting constraints on the final state which the system can settle into. When the system runs, the activities of individual units will change in a way which increases the number of these constraints which are satisfied. Thus, connectionist networks are said to function by constraint satisfaction.
A system which functions through constraint satisfaction has a number of desirable characteristics for modelling human cognition. The main one is that it allows a decision to be reached by a consensus of evidence, a reasonable fit between input and memory, rather than requiring an exact match. This is a virtue in any model of human cognition because the nature of the nervous system requires a degree of fault tolerance in the information-processing system. It is also desirable given the nature of the input which the cognitive system has to work with in the real world. For example, consider what happens when a person listens to one particular speaker in a crowded room. The signals arriving at his/her ear contain the sounds made by the person listened to, with a jumble of sounds from different speakers superimposed on these, and obliterated from time to time by bursts of laughter and other noises. And yet, most of the time, what the listener perceives is words. The signal received bears some relationship to a prototypical representation of the word which the listener perceives but will be far from an exact match. The fact that the person perceives words shows that the word recognition system must be looking for a best fit to the word patterns it has stored rather than for an exact match. This effect has been studied in the laboratory with an experimental paradigm called 'phoneme restoration'. In a study by Warren (1970) the sound /s/ was removed from the word 'legislature' and replaced with a cough. People then listened to a sentence containing the word 'legi<cough>lature' and were asked what they heard. People reported the sentence correctly, adding that there was a cough before or after the word 'legislature'. In other words, the perceptual system does not necessarily give a veridical account of the stimulus; it gives a plausible interpretation of the input, given its knowledge of English words.

Equivalence of 'memory' and 'processing' in connectionist models

In distributed representations the distinction between memory and processing is blurred. Traditional models of cognitive processes often distinguish between 'memory', a store of learnt information, and 'processing', operations which enable the system to interpret incoming information. The processing operations may use information from memory, but the conceptual distinction is clear. Indeed, in many models this is made explicit with separate parts of the model labelled 'memory' and 'processor'.

There is no such distinction in connectionist models. All the information which the network has – its memory – is stored in the weights of the connections between units. All the processing that the network can do is determined by the same set of weights.
At one extreme it seems clear that some sorts of knowledge can be acquired immediately, without interfering with other information. A novel piece of information which one finds interesting or amusing, in a domain where one already has sufficient knowledge to understand its significance, is likely to be remembered after a single presentation. Moreover, it can be retrieved as a specific item of information in future, independent of any other facts in the database. It seems unlikely that such acquisition is accompanied by the loss of any other information. Quick, cost-free addition of new information to existing databases characterises certain sorts of human knowledge acquisition. That connectionist models can perform one trial learning is demonstrated in a model of the role of the hippocampus in episodic memory formation.

At the other extreme there are many areas of knowledge acquisition, such as learning to play tennis or learning to talk, where acquisition of new knowledge is gradual, and accompanied by the modification or loss of previous patterns. As a person’s tennis serve improves or he/she learns to pronounce the language correctly, the brain has to lose some aspects of old response patterns because they were inaccurate. Later it is difficult to recall when a specific piece of information was added to the database. Such a pattern, where new information is inextricably interwoven with old memories, occurs naturally within a distributed system, but not with a localist storage system. Distinctions between different sorts of knowledge representation occur in many models of the human cognitive system.

C.2.3 Selective history of early connectionism

It may seem that connectionism burst upon cognitive science in 1986 with the publication of Parallel Distributed Processing by McClelland and Rumelhart. Although this book had more impact than any other publication in the history of connectionism, it was preceded by decades of research into the computational abilities of networks of simple computing elements working in parallel.

McCulloch and Pitts (1943)

A key event in the history of connectionism was the publication in 1943 of a paper called A logical calculus of the ideas immanent in nervous activity by McCulloch and Pitts. They demonstrated that a network of simple computing units, operating in parallel, could perform logical operations. A crucial aspect of their modelling was that the properties of the computing units were based on those of the neuron (as they were believed to be at the time the paper was written). This approach
has been followed (up to a point) by contemporary connectionism. It was not followed by the mainstream of cognitive function modellers, for whom the aim was to model what the cognitive system achieved, not how it did it.

McCulloch and Pitts' 'neurons' had three properties: They were binary state devices, either 'On' or 'Off'; they summed input, either excitatory or inhibitory, from other neurons which were 'On'; and they had a threshold. Computation took place in a succession of time slices. In each time slice every neuron summed its excitatory inputs. If these exceeded its threshold, it went into the 'On' state for the next time slice, otherwise it remained in the 'Off' state. A single inhibitory input forced it into the 'Off' state for the next time period.

One difference between this and most later work is that McCulloch and Pitts modelled the computational unit as an all-or-none device. Later work usually allowed a continuous value for the state of each unit. This reflects the belief that nerve cells transmit a continuous range of information (for example, by firing rate), rather than just the single fact of being active or inactive. A second difference is that McCulloch and Pitts used unmodifiable connections between units. Later work was often concerned with learning, and the modification of connections was usually the way that learning was implemented. A third difference is that McCulloch and Pitts demonstrated the ability of such nets to perform logical operations. Later work has tended to concentrate on the ability of neuron-like nets to solve problems by statistical combination of evidence rather than by the application of logical relationships.

Hebb (1949)

In 1949 Hebb published The Organisation of Behaviour. The lasting importance of this book came from the proposal of a specific synaptic change which might underlie learning. This is known as the Hebb synapse. Hebb put forward the case for basing models of cognitive function on what was known about the physiology of the brain. Hebb wrote: "The first object of this book is to present a theory of behaviour for psychologists; but another is to seek a common ground with the anatomist, physiologist and neurologist, ... to make it possible for them to contribute to (psychological) theory."

Hebb thought it would help psychological theorising if there was an input from neurophysiology. He contrasted this approach with the deliberate attempt to get neurophysiology out of psychological theories, represented in particular at that time by B.F. Skinner. For Skinner, the aim of a
psychological model was to represent the rules which would allow behaviour to be predicted from knowledge of the organism’s previous experience and the stimuli now facing it. The mechanism by which this behaviour was generated was, to Skinner, irrelevant. One view of those who maintained that neurophysiological plausibility was irrelevant to psychological theorising, was (and still is) that the gap between knowledge of what brain structure is and how it produces behaviour is so wide, that to try and make psychological theories neuronally plausible is at best unhelpful, and at worst positively misleading. Crick has warned about the dangers of trying to model the brain while ignoring basic facts about its structure (Crick, 1989; Crick and Asanuma, 1986). Attempts in the mid-1950’s to test Hebb’s hypotheses about the formation of cell assemblies by the creation of interneuronal synaptic connections found that the more constraints based on neurophysiology that were built into their model, the better it worked.

Hebb was interested in the changes which took place in the nervous system during learning. He proposed that learning takes place by synaptic modification, that is, by modifying the strength of the connections between computing elements. This became one of the central ideas of connectionism.

For Hebb the result of forming these connections was not simply to form direct connections between stimulus and response. It was to create interconnections between a large, diffuse set of cells, in different parts of the brain, which he called ‘cell assemblies’. Hebb believed that the firing of a sequence of cell assemblies would correspond to a thought process.

Rosenblatt (1958)

The first neural network models which are clearly recognisable as forerunners of contemporary models are those investigated by Rosenblatt (1958, 1962). He called them perceptrons. This work is similar to McCulloch and Pitts’ in that it explored the computational ability of networks of simple computing elements operating in parallel. However, Rosenblatt thought that showing how such networks could perform complex logical operations was inappropriate for understanding information processing by the brain. His work was based on three assumptions which contrast with those of McCulloch and Pitts:

a) Since the connections in the brain do not initially allow useful computations, order will be brought about by changes due to experience through learning.

b) Since individual components of the system are unreliable, computation must be by probabilistic combination of evidence, rather than by precise logical operations.
c) The task for a computational unit in the brain is to discover similarities and differences between stimuli (or classification), rather than to perform logical operations.

Rosenblatt found that although perceptrons could learn to classify large numbers of stimuli, they were poor at extrapolating beyond the examples they had learnt during training to other members of the class which they had not seen. Learning the examples did not lead to discovery of the concept from which the examples were derived. They were good at learning responses to specific stimuli, but poor at performing tasks which required the discovery of some higher order abstraction based on them.

By showing that there were other, apparently simple, features of the visual world which perceptrons could not compute, as well as showing that they would require implausibly large numbers of connections or amounts of time to learn some functions that they could compute, Minsky and Papert (1969) concluded that there was no point in pursuing perceptrons as a basis for perception. This attack was very effective at the time, and both interest and funding for network approaches to cognitive function dwindled. Artificial Intelligence, which tries to find algorithms which can solve cognitive problems without any attempt to use computational elements which are neuron-like, became more popular.

Rosenblatt thought that multi-layer nets containing hidden units would solve problems that single layer perceptrons failed on. The problem was that no learning algorithm for training them was known. That is, with more than one layer of modifiable weights in the system no one knew of a way to decide which weights to adjust on a given learning trial. The discovery that the back-propagation algorithm solved this problem led to the great leap forward in the problem solving power of neural nets in the mid-1980s. Minsky and Papert were aware of the greater power of multi-layer perceptrons but were pessimistic about finding an effective way to train them.

Hinton and Anderson (1981)

Despite the depressing effect of Minsky and Papert’s book, work continued on neural nets in the 1970s. Unsupervised self-organising networks operating by competition to perform feature analysis were described by Von der Malsburg (1973). Formal analyses of the mathematical foundations of neural networks were being developed by Grossberg (1976) and Arbib (for example Amari and Arbib, 1977). The basic properties of pattern association networks were analysed in simple models (Willshaw, 1981). The number of distributed patterns which could be associated
together in models with binary synapses and binary neurons was analysed. The dependence of this number on the sparseness of the distributed representation was demonstrated. Many of the properties of autoassociation networks with recurrent connections and their ability to store memories and later retrieve the whole memory from a part were described and analysed (Kohonen, 1984).

One of the first sets of papers to spell out the implications of these developments for understanding cognitive functions such as memory and perception was Parallel Models of Associative Memory edited by Hinton and Anderson (1981). Pattern associators, autoassociators, competitive networks and self-organising maps were all described. It was a major step in making evident the relevance and promise of network models for understanding not only brain function, but also cognitive function.

C.2.4 Architecture and properties of a pattern associator

During training a pattern associator is presented with pairs of patterns. If learning is successful then the network will subsequently recall one of the patterns at output when the other is presented at input. After training, a pattern associator can also respond to novel inputs, generalising from its experience with similar patterns. Pattern associators are tolerant of noisy input and resistant to internal damage. They are capable of extracting the central tendency or prototype from a set of similar examples.

Architecture and operation of a pattern associator

A fundamental task for the nervous system is to discover the structure of the world by finding and learning to correlate or associate one stimulus with another. For example, one stimulus might be the taste of chocolate and the other its appearance. Initially there is no connection between these two for a child. After the association between the patterns of neural firing caused by the two stimuli has been made, the sight of chocolate can recall responses originally associated with its taste. If the taste of chocolate caused salivation, then the sight of chocolate, which was initially neutral, could produce the same response. This form of pattern association is called Pavlovian conditioning in experimental psychology. However, similar mechanisms underlie far more than the control of autonomic nervous system responses like salivation. Learning that the letter string YACHT is pronounced /y/ /o/ /t/, that a particular person has blonde hair, or how far one has to turn
the steering wheel of a car and how fast one can go around a corner are all examples of pattern association.

The process of pattern association in a single layer connectionist network (a system with a single layer of modifiable connections between input and output units) is shown in Figure C.3. Pattern $P_2$ (1 0 1 0 1 0) is presented to the input units. $P_1$ (1 1 0 0) is the pattern which the network has to learn to produce on the output units in response to the input of $P_2$. The connection between input unit $j$ and the output unit $i$ is represented by the line labelled $w_{ij}$. Pattern association takes place by modifying the strength of the connections between input units and output units.

During the learning phase the two patterns which are to be associated, $P_1$ and $P_2$, are presented to the network simultaneously. Hebbian learning can be implemented in the network in the upper part of Figure C.3 by applying the following rule: If there is activity on input axon $j$ when neuron $i$ is active, then the strength of the connection, $w_{ij}$, between axon $j$ and dendrite $i$ is increased (neuron $i$ will be active if the $i^{th}$ axon in $P_1$ is ‘On’). For the network in the lower part of Figure C.3 Hebbian

![Figure C.3 A pattern association network in the nervous system (a), and in conventional connectionist format (b) (McLeod et al., 1998)](image-url)
learning would be implemented, following the simultaneous presentation of \( P_2 \) at input and \( P_1 \) at output, by strengthening the connections between those units which are ‘On’ in both the input pattern and the output pattern.

The Hebb rule for weight change \( (\Delta w_{ij}) \) can be expressed formally as:

\[
\Delta w_{ij} = \varepsilon a_i a_j
\]

where \( \varepsilon \) is a learning rate constant which specifies how much a synapse alters on any one pairing of \( P_1 \) and \( P_2 \), \( a_i \) is the activity of element \( i \) in \( P_1 \), and \( a_j \) is the activity of element \( j \) in \( P_2 \).

The Hebb rule is expressed in multiplicative form to reflect the idea that, for a synapse to increase in strength, both pre-synaptic activity (from \( P_2 \)) and post-synaptic activity (from \( P_1 \)) must be present. Application of the Hebb rule strengthens every connection between an axon and a dendrite where both are active. At recall that connection is then used to activate the dendrite when the axon in the recall cue is active.

When learning has taken place, the effectiveness of the memory created is tested by presenting a recall pattern \( (P_R = P_2) \) to the network, on its own, on the axons which originally carried \( P_2 \). The consequence should be the recall of pattern \( P_1 \) at the output. The net input to output neuron \( i \) is:

\[
\text{netinput}_i = \sum_j a_j w_{ij}
\]

where \( \sum_j \) indicates that the sum is over all the input axons indexed by \( j \). Output neurons could have a binary threshold activation function.

**Properties of a pattern associator**

During recall, pattern associators generalise. That is, if a recall cue is similar to a pattern that has been learnt already, a pattern associator will produce a similar response to the new pattern as it would to the old. Small differences in netinput, for similar input patterns can be removed by a threshold function at the output. In the world faced by real biological systems, recall cues are rarely identical to the patterns experienced during learning. Therefore, a mechanism which automatically generalises across slight differences in input patterns has obvious adaptive value.
Even if some of the synapses on neuron $i$ are damaged after learning, $\text{netinput}_i$, following the presentation of a recall cue, may still be a good approximation to the correct value. Provided the pattern carrying the recall cue consists of a reasonably large number of axons the correlation will not be greatly affected by a few missing items. After passing through the binary threshold activation function, the result may well be correct recall. The same result is achieved if some of the input axons carrying the recall cue are lost or damaged. Since real nervous systems are continually losing cells, fault tolerance is of great adaptive value. The properties of generalisation and graceful degradation are only achieved if the representations are distributed.

Recall is very fast in a real neuronal network. The input firings which represent the recall cue $P_R$ are applied simultaneously to the synapses, so $\text{netinput}$, can be accumulated in one or two time constants of the dendrite (for example 10-20 msec). If the threshold of the cell is exceeded, it fires. Thus, in no more than a few tens of milliseconds all the output neurons of the pattern associator, which will be turned ‘On’ by a particular input pattern, will be firing. The time taken to switch the output neurons ‘On’ will be largely independent of the number of axons or dendrites in the pattern associator. These types of neuronal networks operate fast in the brain since they perform parallel processing. This is very different from a conventional digital computer. Computing the output of a connectionist pattern associator involves successive multiplication and addition operations for the net input to each output unit. The time to compute the output pattern would increase in proportion to the product of the number of axons and the number of dendrites.

Learning is also potentially fast in a pattern associator. A single pairing of $P_1$ and $P_2$ could, in principle, enable the association to be learnt. There is no need to repeat the pairing over many trials in order to discover the appropriate mapping. This is important for biological systems. A single co-occurrence of two events may provide the only opportunity to learn something which could have life-saving consequences. Although repeated pairing with small variations of the vectors produces the useful properties of prototype extraction and noise reduction, the properties of generalisation and graceful degradation can be obtained with just one pairing.

Independent events can be stored on the same connections of a distributed memory system. However, it is obvious that there will be limitations on the number that can be stored without interference. Interference between responses to different but similar input patterns is the basis of many important properties of distributed memories and allows generalisation, noise reduction and prototype extraction. Interference is a major property of human memory, while generalisation underlies many cognitive processes. One reason that interference is tolerated in biological memory

348
is presumably that the ability to generalise between stimuli is more useful than 100% accurate memory of specific past events.

C.2.5 Architecture and properties of an autoassociator

An autoassociator reproduces at the output units the same pattern which was presented at input. It has many basic properties in common with the pattern associator taught with the Hebb rule. It can store independent memories on the same set of connections and it has the desirable property of ‘cleaning up’ incomplete or noisy inputs. When individual experiences, drawn from a set of different categories, are stored in an autoassociator memory, prototypical instances of each category are formed automatically. This offers a possible solution to the problem of how humans succeed in categorising the world without explicit teaching.

Architecture and operation of an autoassociator

The difference between the structure of an autoassociator and that of a pattern associator is that the output line of each unit is connected back to the dendrites of the other units. These are called recurrent connections. They are represented by the loop back from the output of each unit which connects to the dendrites of all the other units as shown in Figure C.4. At each meeting of recurrent feedback from one unit and the dendrite of another, there is a modifiable connection.

![Figure C.4 An eight unit autoassociator (McClelland and Rumelhart, 1985)](image)

The net input to unit $i$ consists of the external input ($\text{extinput}_i$) and the internal input generated by feedback from other units within the autoassociator:

$$\text{netinput}_i = \text{extinput}_i + \sum_j a_{ij} w_{ij}$$
where \( a_j \) is the activity of unit \( j \) (in this case \( j \) indexes the units other than \( i \)) and \( w_{ij} \) is the strength of the connection between the recurrent input from unit \( j \) and the dendrite of unit \( i \).

The external input produces an activity level in each unit determined by an activation function. Via the internal feedback connections, this activity, weighted by the strength of the appropriate connection, then starts to produce an internal input to all the other units. The net input to a unit is now the sum of the external and internal inputs. This produces a new level of activity in the unit, which feeds back to all the other units, changing their net input and activity level, and so on. There is a danger that activity in a system with positive feedback can continue to grow. The use of a non-linear activation function such as the sigmoid function ensures that the activity of each unit cannot grow beyond a fixed maximum value. The autoassociator is allowed to run for a number of cycles until it reaches a steady state where there is no further change in the internal input.

The aim of the autoassociator to reproduce at output the pattern presented at input is achieved during the learning phase by changing the weights so that the internal input to each unit matches the external input. This can be done by calculating the difference between the internal and external inputs and changing the weights of the connections in the direction which will reduce the difference. This is an example of the application of the Delta rule. The difference between the external and internal (intinput,) inputs to unit \( i \) is \( \delta_i \):

\[
\delta_i = \text{extinput}_i - \text{intinput}_i
\]

The rule for the weight change (\( \Delta w_{ij} \)) in the connection between unit \( i \) and the recurrent input from unit \( j \) is:

\[
\Delta w_{ij} = \varepsilon \delta_i a_j
\]

where \( \varepsilon \) is a constant which determines how large the weight change is on any individual trial. \( \delta_i \) is the error on unit \( i \), and \( a_j \) is the activity of unit \( j \). The inclusion of \( a_j \) in the weight change rule is a way of apportioning blame for \( \delta_i \). \( \delta_i \) represents the failure of the internal input to unit \( i \), \( \Sigma_j a_j w_{ij} \), to match the external input. By making the weight change on a connection proportional to \( a_j \), the activity of the unit sending input along that connection, the changes are concentrated on those connections where they will have most effect on the internal input to \( i \).
It is possible to train an autoassociator with the Hebb rule or a pattern associator with the Delta rule. The Hebb rule operates by strengthening connections between units which are both active. The Delta rule starts by defining a desired state of affairs and operates by strengthening connections which would lead to a reduction in the difference between the desired and the actual state of affairs.

Properties of an autoassociator

An autoassociator can complete a fragment of a pattern previously learnt. A dramatic example of the ability of an autoassociator to recall complete memories from partial or noisy cues is shown in Figure C.5. Therefore, an autoassociator memory stage might be included in a processing system. It will not just reproduce the input, but if the external input is incomplete or noisy, the internal input can clean it up before passing it on to the next stage of processing. Given that the brain is a noisy environment where signals are often distorted, either by random cell firing or by random cell death, this has obvious benefits. It is also the case that many signals reaching the brain are noisy versions of a prototype. Every example of a spoken word which one hears is slightly different because of the accent of the speaker, or the context it appears in, or the background noises which are superimposed upon it. Generally, the distortions are not interesting in themselves – it is the prototype which the distortion is coming from which one wishes to identify.

![Input Output](image)

**Figure C.5** Recall of a complete image by an autoassociator given a noisy or partial clue (Hertz et al., 1991)

Despite its usefulness in some situations, there is a cost to these cleaning up operations. The ‘incomplete’ or noisy’ input may be neither – it may be a new signal, conveying some information.
different from stored patterns. The autoassociator cannot know which is the case. An external input, which has not been presented before will be altered to look more like whichever of its previously learnt signals the new one is most similar to. It does force any part of the system beyond the autoassociator to see the present through a distorting filter of past experiences. This is for sure a human characteristic. Everyone will recognise in other people (if not in themselves) a tendency to interpret novel experiences, not as they really are, but through a filter of preconceptions, determined by past experiences.

An autoassociator memory can come to represent the central tendency of a set of related examples without experiencing the prototype itself. Since people appear to do this, it is an important property of a memory system to simulate. The autoassociator offers a solution to one of the central mysteries of cognition. With the right memory architecture, mere experience is enough to ensure that appropriate categorisation of stimulus events will take place without a homunculus to tell the system what would be an appropriate classification. This ability is not restricted to autoassociators. Competitive nets also perform untutored categorisation of inputs.

C.2.6 Training a multi-layer network with an error signal – hidden units and back-propagation

A powerful technique for training networks is to use a measure of error to change the weights. This discovery that networks with hidden units could be trained with the back-propagation algorithm is probably the single most important factor in the growth of interest in connectionist networks which began in cognitive science in the mid-1980s.

The importance and utility of the back-propagation algorithm can be appreciated by considering a number of separate modelling issues:

a) Using the error of the output to adjust the weights is a powerful way to teach a network.

   The technique of error gradient descent is particularly effective, guaranteeing to find a set of weights to perform the task (that is to produce the correct output pattern in response to all input patterns) if one exists.

b) With single layer networks, a set of weights which will produce zero error only exists if the input-output mappings to be learnt are linearly separable. However, many problems of interest in cognitive science do not satisfy this condition.

c) Multi-layer networks can solve problems which are not linearly separable by forming a novel representation of the input patterns on the hidden units. The ability to effectively
reorganise a problem into a form which is soluble by employing hidden units, makes multi-
layer networks extremely effective problem solving devices.

d) Error gradient descent learning requires the error on all units in a network to be known.
The error on the output units of a multi-layer network is known but the error on the hidden
units is not.
e) The back-propagation algorithm suggests a plausible way to estimate a value for the error on
each of the hidden units given the error on the output units.

Error gradient descent requires the units in the network to have a differentiable and therefore
continuous activation function. A favoured activation function in connectionism, the logistic
function, meets this requirement.

**Back-propagation and biological plausibility**

Back-propagation requires connections to carry signals in two directions – activity level in one
direction and error in the other. When it was first proposed as a learning algorithm it could not be
verified to be biologically inspired. It has seemed implausible that back-propagation could be
implemented in the brain because axons are basically unidirectional transmitters of information.
However, the technique is so effective that it has been widely used as a way of training
connectionist networks in the simulation of cognitive processes. The use of back-propagation has
led to polarisation of the connectionist modelling community depending on whether biological
plausibility or effective simulation is seen as more important.

A response favoured by some connectionists using back-propagation to train their networks is that
they are studying what networks learn rather than how they learn it. Their assumption is that
although back-propagation may not be used by the nervous system, the representations it forms are
similar to those which would be formed by any gradient descent learning algorithm. Therefore, the
behaviour of any network trained by a gradient descent learning algorithm would be similar
irrespective of the way gradient descent is achieved. Currently researchers do not know enough
about neurons to know how the learning of higher cognitive functions is achieved in the brain.
Back-propagation is a technique which may turn out to be biologically implausible. Similarly,
other connectionist learning techniques will require modification as more is discovered about
synaptic modifiability.
However, some neural network modellers do not use back-propagation because it seems implausible to assume it to be implemented in the brain. First, the algorithm requires a single hidden unit to receive information from all the neurons to which it is connected concerning their error, and to know the relevant strength of every synapse involved, to compute an error term. Second, for back-propagation networks to function correctly, the number of neurons in the hidden layer is crucial. If there are too many, the network fails to generalise, and if there are too few, the network fails to learn correctly. There is no evidence (as yet) about how the brain could set the number of hidden units available to solve a cognitive problem to an appropriate value. Modellers who find these arguments persuasive use learning rules such as the Hebb rule. The Hebb rule is an example of a local learning rule – so-called because the information required to implement it is believed to be present at the synapse where the learning takes place. Thus, the Hebb rule is considered more biologically plausible than back-propagation.

C.2.7 Competitive networks

When an input pattern is presented to a competitive network the output units compete with each other to determine which has the largest response. This unit is the ‘winner’ for that input pattern. The connections to the winning output unit from the input units which were active in that pattern are strengthened and those from input units which were inactive are weakened. When this learning algorithm has been applied to the different winning units following a range of input patterns, the network will come to categorise input patterns into groups, with one output unit firing in response to each.

Hebbian and Delta learning together with back-propagation constitute ways to teach a network to produce a specific output in response to a specific input. The desired response of the network acts as an explicit teacher signal. During training, any discrepancy between the current output and the desired output is used to adjust the weights until the desired output is obtained. These methods are examples of supervised learning since the response which the network is required to learn is presented to the network during training. However, useful knowledge can also be acquired by networks when there is no pre-determined target for each output unit. This is called unsupervised learning. A competitive network learns to categorise input patterns into related sets, with one output unit firing for each set.
Architecture and operation of a competitive network

Competitive learning can be demonstrated with the simple network shown in Figure C.6. Three input units are connected in a fully feed-forward fashion to two output units which have inhibitory connections with each other.

![Figure C.6 A simple competitive learning network with inhibitory connections in conventional connectionist format (a), and in the nervous system without inhibitory connections (b) (McLeod et al., 1998)](image)

Excitation propagates from the input units to each of the output units. These compete with each other via the inhibitory connections until only one remains active and becomes the 'winner'. In some competitive networks, there is a gradation of success for different units, not just one winner, but the principles are similar. The strengths of connections to the winning output unit from active input units are then increased and those from inactive input units are decreased. Connections to units which did not win the competition are not altered. This procedure does not involve an external teaching signal. Learning is determined by which output unit produces the biggest response to the input signal, the outcome resulting from the interaction between the input signal and the weights on each output unit. The weights are set by the prior learning of the network, not by an explicit external teacher. Thus, competitive learning can be divided into three phases: excitation, competition and weight adjustment.

Competitive learning is biologically plausible because it uses a local learning rule. That is, the information required to change the weight of a connection is available at the axon and the dendrite on either side of the connection. Competitive learning is also of interest in brain computation because it can perform feature analyses without an explicit teacher. The only difference between this learning paradigm and the pattern associator is the absence of the external input driving the output neurons because there is no external teaching input in competitive learning. The output neurons are activated in the usual way by the pattern of activity on the input axons – netinput\(_i\) = \(\sum_j a_j w_{ij}\). They then compete with each other (in the brain, for example using inhibitory
interneurons) until the most strongly activated neuron is the only one firing. Hebbian learning is then applied – synapses between input lines which are active and output neurons which are active are strengthened. Since the only output neuron which is active is the winning unit, the only synapses which are strengthened are those between the active input lines and the winning output unit. This achieves the basic goal of competitive learning ensuring that the winning output neuron is more likely to win if the same or a similar pattern is presented in future.

Properties of a competitive network

Competitive networks are a feature of many real brain circuits. This is not surprising as they have a number of potentially useful properties. First, they can remove redundancy from a set of inputs by allocating a single output neuron to represent a set of inputs which co-occur. This reduces the number of active units required to represent an input pattern without a loss of information. This can be a helpful pre-processing operation in networks which require a sparse input to operate efficiently. Second, they can produce outputs for different input patterns which are less correlated with each other than the inputs were. In the limit, they can turn a set of partly correlated patterns at input into a set of orthogonal or uncorrelated patterns at output. This is a useful function because pattern associators can work more efficiently if the input patterns they have to learn are not highly correlated. The efficiency of a pattern associator can thus be improved by allowing a competitive network to act as a pre-processor on its input. Such a network can solve problems which would be too hard for a single layer network without introducing the complexities associated with teaching a multi-layer network.

Orthogonalisation and categorisation might seem to be entirely different processes. The first involves producing outputs which are less similar to each other than the input patterns which generated them. The second involves producing the same output to inputs which are not identical. However, competitive networks can perform both operations. By allocating separate output units to somewhat related input patterns, a competitive network can produce a set of outputs which are less correlated than the inputs – orthogonalisation. By allocating the same output unit to more related patterns a competitive network can produce outputs which are more correlated than the inputs – categorisation. The balance between the number of output units and the number of patterns presented to the network will determine which process takes place.

A common competitive architecture is to have a succession of layers, each operating as a competitive net. After the initial input layer, the units within a layer are collected into competitive
clusters. Within each cluster only one unit, the winner, will be active. Any unit within a layer can have an excitatory connection to any unit in the next layer so the input to each successive layer will be the winner from each of the clusters in the previous layer. A series of competitive transformations will take place, with the potential for discovering successively higher order patterns in the original input. An example of what can be achieved by a multi-layer competitive network is the development of position and view invariant representations of images of faces in the simulation of processing of visual information in the temporal lobe.

Topological organisation is found in many cortical processing areas. That is, similar input features are processed in adjacent cortical regions. A simple modification to the competitive networks described so far enables them to develop topological maps. The modification is to add short-range excitation and long-range inhibition between the units. The effect of this connectivity between units, which need not be modifiable, is to encourage units which are close together to respond to similar features in the input space, and to encourage units which are far apart to respond to different features in the input space. When these response tendencies are present during learning, the feature analysers which are built by modifying the connections from the input to the activated units tend to be similar if they are close together, and different if far apart.

The biological utility of developing topology-preserving feature maps may be that if computation requires neurons with similar types of response to exchange information more than neurons involved in different computations, then the total length of the connections between the neurons is minimised if the neurons which need to exchange information are close together. Minimising the total connection length between neurons helps to keep the size of the brain small. Placing neurons close to each other which need to exchange information, or which need to receive information from the same source, or which need to project towards the same destination, may also help to minimise the complexity of the rules required to specify cortical connectivity.

C.2.8 Recurrent networks

In a recurrent network, output from later layers feeds back to provide new input for earlier layers. Such networks can produce sequences of output following a single initial input or predict the next input in a sequence. They can also form attractor networks in which the output in response to an input changes with time.
Much of human activity involves the coordination of motor sequences to achieve goals. For example, the production of a word involves the utterance of a sequence of phonemes in the appropriate order. If the sequence is lost, the word changes its identity or becomes nonsensical.

**Simple recurrent networks (SRNs)**

A simple recurrent network (SRN) contains connections from the hidden units to a set of context units. These context units store the hidden unit activities for one time step, and then feed them back to the hidden units on the next time step. They are the equivalent of state units, storing the network’s memory of the state of the network on the previous time step. Since the hidden units have an input which includes a record of their prior activity, they are able to carry out tasks which extend over time. This sort of network is not merely a tape-recording of the past. The hidden units continue to recycle information over multiple time steps. The input to the hidden units at time $t + 1$ includes information for $t$ and $t + 1$. At time $t + 2$ the input includes information from $t + 2$, $t + 1$ and $t$. And so on. If there are sequential dependencies in the training data, an SRN can discover them. Anticipation plays a key role in early learning, so learning to predict is an important aspect of cognition.

![Diagram of a simple recurrent network](image)

**Figure C.7** A simple recurrent network (SRN) (McLeod et al., 1998)

In Figure C.7 the activities of the context units are direct copies of the hidden unit activities on the previous time step. Connections from the hidden units to the context units are fixed while connections from the context units to the hidden units are adjustable through the learning procedure, in the same manner as all the other connections in the network. Context units can be thought of as providing the network with a dynamic memory. Identical input signals can be treated differently depending on the current status of the context.
Recurrent networks are trained in the same way as feed-forward networks. First, an input pattern results in activity being propagated through the network to the output units. The actual pattern of activity at output is then compared to the desired output (usually the next pattern in the sequence) and the discrepancy between the two is used to drive a back-propagation learning algorithm to adapt the weights in the network. If the desired output is the next pattern in the sequence, the network is being trained to predict the next input. Weight adaptation occurs in just the same way as in a feed-forward network, except that the connections from the hidden units to the context units are never changed since their function is to establish copies of the hidden unit activities on the context units.

It is possible to create different types of recurrent network architectures. For example, an autoassociator contains recurrent connections between all the units in the network. Another type of recurrent network, referred to as a Jordan net, uses the activity of the output of the network (rather than the activity of the hidden units) to feed back to the context units which themselves may have self-recurrent connections. These differences in network architecture result in differences in computational properties.

**Attractors**

The output of a simple recurrent network changes over time in a well-defined way so it can learn to produce or predict sequences. It is also possible to devise networks where the change of the output over time causes the network to settle into one of several states depending on the input. This is called an attractor network. The set of possible states into which the network can settle are the attractors.

![Figure C.8](image.png)

**Figure C.8** An arbitrary mapping problem with input patterns (+s and −s) and attractors (dots) (a), solved by a recurrent network creating two basins separated by the curved line (b)

(McLeod et al., 1998)
Attractor networks have a number of properties which make them very useful in the simulation of cognitive processes. First, they produce a simulation of reaction time. A stable output (which would be required for a response) will require the network to settle into one of the attractor states. The ‘reaction time’ of the network is the length of time (in terms of the number of cycles) it takes to reach a stable state. The time it takes to do this will depend on the shape of the so-called ‘basin of attraction’ and the position within the basin where the initial input landed. Second, attractor networks are relatively immune to noisy input. Provided an input falls within a basin, the exact position does not matter. The response will eventually be the same – the state corresponding to the attractor. Third attractor networks allow an arbitrary mapping between input and output. Suppose a network has to learn a mapping where an arbitrary set of inputs, some of which are close to (0.1 0.1), must produce the output (0.9 0.9) and vice versa. Such a desired mapping is illustrated in Figure C.8. A recurrent network can learn to produce a boundary between two attractor basins so that the appropriate mapping will be achieved. The ability to do this is important because many cognitive tasks involve arbitrary mappings.
C.3 Applications of Connectionist Models

C.3.1 Reading aloud

Many traditional box-and-arrow models of the information processing involved in skilled reading have two independent mechanisms which can lead to the pronunciation of a letter string, the rule based and the lexical or word based, represented as two separate pathways (hence ‘2-route’ models). An example of such a model is shown in Figure C.9. The lexical route uses a store of information about specific words to generate the correct pronunciation (directly, or via semantics); the non-lexical route consists of a set of pronunciation rules. Evidence consistent with this sort of model comes from neuropsychology. There is a striking contrast between the pattern of reading errors made by two groups of patients who have become dyslexic as a result of brain damage acquired after they have learnt to read. Phonological dyslexics read words without difficulty, but cannot produce a pronunciation for non-words. A second group, called surface dyslexics, pronounces regular words and non-words correctly but makes errors on irregular words, tending to regularise them. This contrast is easily explained by 2-route models.

![Figure C.9 A typical ‘2-route’ model of reading aloud (McLeod et al., 1998)](image)

Although non-connectionist models of reading aloud differ in many details, there is general agreement on two issues. First, the reality of a privileged status for words is not in doubt – all models contain a lexicon, a store of specific information about particular words with which the
reader is familiar. To allow pronunciation of novel letter strings, there is another mechanism, such as a set of general rules of pronunciation.

Connectionist models of reading challenge both these assumptions. There is no place in a distributed connectionist model for a lexicon where specific information about particular inputs is stored, independent from other information. Information is distributed across the storage space, superimposed on the information associated with other inputs. Similarly, there is no distinction between specific information and general rules. There is only one kind of knowledge – the weights of the connections which the model has acquired as a result of its experiences during training, all stored in a common network. In order to build a relative complete model of reading aloud, it is important to note that skilled readers do not just know how to read words; they also know what they mean. A model which has no representation of the semantics of the words it has learnt, cannot make use of semantic knowledge to help is tasks like lexical decision.

A model developed by Plaut et al. (1996) has shown that it is possible for a connectionist model of a real, complex cognitive process to produce a precise quantitative fit to data produced in experiments with normal subjects. The success of the model with attractor structure suggests that this approach might also work in a realistic noisy environment. The fact that a distributed model can work shows that neither localist information storage nor knowledge represented as an explicit set of rules is a requirement for modelling cognitive processes. However, the model suggests that localist and distributed information storage do not necessarily have to be seen as polar opposites. The attractor basins which emerge as the dynamic state of the model interacts with specific inputs seem to have some of the characteristics of lexical units. Furthermore, by building componential attractors for the individual parts of regular words and non-componential attractors for exception words, non-words are given a regular pronunciation (like MAVE) and exception words are pronounced correctly (like HAVE). The distinction between componential and non-componential attractors seems to offer a way in which, what might be regarded as rules of pronunciation, could be instantiated.

The success of the simulation by Plaut et al. (1996) highlights one problem of explicit computational models. Some input coding system must be chosen to get the model to run. An inappropriate form of input coding can prevent the model from simulating observed normal behaviour, as was the case with the inability of a model by Seidenberg and McClelland (1989) to simulate reading aloud. However, the failure was not of the connectionist principles which were
under investigation. It was of an arbitrary component of the model which was irrelevant to the central issue of whether the connectionist approach to cognition is a useful one.

C.3.2 Language acquisition

Summary

A crucial aspect of connectionist modelling, which distinguishes it from much of traditional cognitive modelling, is that it models learning. A connectionist model is not just a model of a skill, but also a model of how the skill is acquired. To evaluate their success as models of learning they can be tested against developmental data.

The connectionist approach to language acquisition offers a direct challenge to much received wisdom in cognitive psychology. Many traditional models assume that language learning involves the acquisition of rules. For example, over-regularisation errors occurring in children’s early attempts to produce the past tense (saying ‘eated’ instead of ‘ate’) are seen (not surprisingly) as evidence for the possession of the rule ‘to form a past tense add –ed to the verb stem’. The fact that children can make the subject and verb agree in a sentence they have never produced or heard before is assumed to reflect the possession of a general rule about the relation between subject and verb which can be applied in any sentence.

Connectionist models do not have explicit declarative rules like ‘to form a past tense add –ed to the verb stem’. Neither is there a place for a specific memory to store the list of exceptions which would be required to override application of this rule to exception verbs. Knowledge about all verbs, regular and irregular, is stored in the same general matrix of information. So a challenge for connectionism is to show how a pattern of behaviour which appears to require a database of rules and exceptions can be produced without one.

Developmentalists try to characterise the nature of the start state which allows the acquisition of skilled behaviour. Connectionist modelling allows a theory about the mechanisms present in the start state to be implemented. For example, the acquisition of inflectional morphology and syntactic rules, show that a relatively unstructured start state can support the learning of a complex skill. However, in both cases the nature of the training environment to which the unstructured start state is exposed is crucial. Behaviour which mimics that of children will only emerge under certain conditions.
A common view in developmental psychology is that a sudden change in performance reflects the acquisition of a new cognitive mechanism. However, modelling of the vocabulary spurt in young children, shows how a system with a constant architecture and learning rule can, nevertheless, produce a sudden change in behaviour. Connectionist modelling shows that discontinuous performance does not necessarily imply a discontinuity in the underlying mechanisms.

Learning the English past tense (Inflectional morphology)

Young children add the past tense suffix ‘ed’ to form the past tense of irregular verbs, producing errors like ‘goed’. Since this would have been appropriate were the verb is regular, such errors are known as over-regularisations. These have been cited as evidence that in learning a language, children are acquiring a system of rules. Over-regularisation errors are assumed to result from the inappropriate application of a rule. They cannot be explained in terms of imitation. The child is not exposed to over-regularisation errors in the adult language. The errors are the child’s own creation. Children would be expected to make errors as they acquire a skill. However, over-regularisation errors are surprising because they often occur after children have succeeded in producing the correct past tense of the verb. For example, children may correctly produce the past tense form ‘went’ early in their third year and then produce ‘goed’ at the beginning of their fourth. Developmentalists have sought an explanation for this U-shaped profile of learning.

A natural interpretation of this pattern of performance is to suggest that early in development, children learn past tense by rote, storing in memory the forms that they hear in the adult language. At a later stage, they recognise the regularities in the inflectional system of English and reorganise their representation of the past tense to include a new device that does the work of adding a suffix. Subsequently they will not need to memorise new forms. During this stage, some of the originally learnt irregular forms may suffer inappropriate generalisation from the addition of the regular suffix. Finally, they sort out which forms cannot be generated with the new rule-based device and these are stored as a list of exceptions.

This approach identifies errors as a consequence of the transition from pure rote learning to partly symbolic rule-governed behaviour. Recovery from error occurs when the representations of the irregular forms are sufficiently strong to resist interference from the rule-governed process. Representational strength is achieved by continued exposure to the language, so infrequent irregular forms will be more susceptible to over-regularisation than frequent forms. Frequent forms will also
recover more quickly from error. The symbolic account identifies the recovery from over-regularisation errors with the consolidation of an associative memory for the representation of irregulars. The associative process operates in parallel and in competition with the rule-governed process. Therefore, the mature state of the adult inflectional system can be characterised as a dual-route device, as shown in Figure C.10.

![Figure C.10 The dual-route model for the English past tense (Pinker and Prince, 1988)](image)

An alternative approach uses a connectionist network. Such a system does not make any initial distinctions between regular and irregular forms, but gradually learns to identify patterns of regularity based on the forms it encounters. In addition, the model does not require any rules to be built into or mature in the system. For example, the model by Plunkett and Marchman (1993) demonstrates how patterns of regularity and irregularity might be represented in a cognitive system without appealing to an innate pre-wiring of the system. The model challenges the orthodoxy that learning a language consists of learning a system of symbolic rules. Of prime importance to developmentalists is the demonstration that a homogenous computational system can learn to perform a complex task which was thought to require a heterogeneous architecture. The model demonstrates that one does not need as much pre-wired, innate structure to learn the past tense as the dual-route approach supposes.

The fact that a multi-layered network can perform the task does not prove that children use the same type of mechanism – only that they might. Evaluation of the plausibility of the model rests on its behavioural predictions. Connectionist models are slaves to their training environments. Their performance reflects the distribution of examples they encounter. For example, current connectionist models predict that the onset of over-regularisation errors, heralding the discovery of the regularities underlying the inflectional system, is closely yoked to the achievement of a critical mass of regular verbs in the child’s vocabulary. A critical mass effect in children acquiring the English past tense has been demonstrated by Plunkett and Marchman (1996).
Connectionist modelling of past tense acquisition shows that inferring dissociations in mechanism from dissociations in behaviour (in this case, performance on regular versus irregular verbs) is hazardous. It is not necessary to invoke a separate symbolic system to explain the processing of regular verbs. Instead, it is possible that both regular and irregular verbs are processed in the same fashion, and perhaps even in the same mechanism. This result reinforces the view that the start state for language learning (at least for inflectional morphology) may not require as much initial structure as has been supposed. Instead, the additional constraints necessary for learning inflectional systems may be found in the learning environment itself – in this case, English verbs and their past tense inflections. One of the strengths of connectionist networks is their ability to extract and represent the patterns of regularity inherent in a structured training environment. In this respect, they act like statistical inference machines. However, they can go beyond mere surface regularities to extract and represent the abstract structure of the input. Their use in modelling development provides the researcher with a tool for examining the trade-off between initial architectural and computational constraints on the one hand and environmental information on the other.

Early lexical development (Vocabulary spurt)

A dramatic increase in vocabulary is observed in many children towards the end of their second year – the so-called ‘vocabulary spurt’. The sudden acceleration in vocabulary size appears to signal the beginning of a new stage of development in children’s language acquisition. Since most of the new words acquired by children at this point are names for objects, the vocabulary spurt has often been interpreted as evidence for the development of a naming insight, though other explanations relating to children’s developing phonological skills or semantic representations have also been proposed. All these explanations assume that the discontinuity is associated with the emergence of a new cognitive mechanism.

Children produce over- and under-extension errors at this stage. The ability to say the word ‘dog’ does not necessarily show that the child has a complete understanding of what it means. The word may be used when referring to a cat – an over-extension error. Alternatively, it may be used only when referring to the family dog – an under-extension error. Such errors are assumed to arise because children base their use of words on the ‘prototype’ of the concept associated with the word. The over-extension error of using the word ‘dog’ to refer to a cat could arise because cats share certain features with the prototypical dog. The under-extension error of a child who fails to use the
word ‘dog’ the first time he/she encounters a dachshund could arise from the fact that it lacks some of the features of the prototypical dog.

Children also appear to understand words that they do not themselves produce. This asymmetry has prompted some researchers to postulate the existence of separate mechanisms for comprehension and production of words. Others maintain that comprehension and production exploit the same underlying knowledge base and that the asymmetry can be explained by the difference of the tasks. For example, production involves recall whereas comprehension involves recognition. It is often difficult to decide between non-computational theories.

A connectionist model of lexical development by Plunkett et al. (1992), shown in Figure C.11, learns to name objects, which produces a vocabulary spurt, over- and under-extension errors, prototype effects and a comprehension-production asymmetry — the network is better at comprehension (generating the correct image at output in response to a name at input) than it is at production (the ability to name an object). The success of this simulation shows that these effects can all be accommodated within a system that maintains the same basic architecture and learning process throughout training. The discontinuities in the performance of the model result from a process of continuous and progressive change in the network, not from the appearance of new mechanisms. The simulation also demonstrates how the connectionist approach can simplify cognitive modelling by showing that an apparently diverse set of behavioural phenomena emerges from the same structure.

![Diagram of the network architecture for lexical development](image)

**Figure C.11** A simplified version of the network architecture for lexical development (McLeod et al., 1998)
When this model was first run it did more than just mimic facts about early lexical development. It predicted new results. For example, the model suggested that children should show a comprehension spurt as well as a production spurt. This was a genuine prediction – it was an emergent property of the network, not one that was ‘designed in’ before the model was built. At that time behavioural data had only identified a production spurt. The comprehension spurt was subsequently confirmed as typical of children’s early lexical development (Goldfield and Reznick, 1992; Fenson et al., 1994).

The model also suggests novel theoretical interpretations of familiar facts. The vocabulary spurt is not due to the triggering or maturation of a new mechanism. The same network architecture and the same learning algorithm are used throughout training. All that changes is the strength of the connections in the network. Initially, the connections in the network are randomised so the system knows nothing about the relationship between objects and labels. Early learning consists in a tentative exploration of the error landscape in an attempt to improve performance on as many input patterns as possible. During the early stages of training, the network is still attempting to make sense of the problem domain to which it is exposed. Solutions for some inputs will be discovered before others. However, as the number of isolated solutions increases, the structure of the problem domain emerges. The network discovers the natural clustering of the objects and the fact that the objects within a cluster tend to possess the same label. The achievement of a critical mass of label-image associations makes the business of learning new associations relatively easy. Hence, a vocabulary spurt is obtained. This explanation illustrates the role of computational models in the understanding of cognitive development. Gopnik and Meltzoff (1987) proposed that the vocabulary spurt emerges from children’s understanding that objects belong in categories and that the discovery of specific semantic domains (such as ‘animals’) helps the child discover which names are associated with which objects. Such a proposal may seen eminently reasonable but without a computational model there is no way of knowing whether it would predict a vocabulary spurt, a steady growth in vocabulary size, or anything else. The fact that the network’s successful categorisation of the objects can be shown to provide the foundations for the vocabulary spurt is the quantitative evidence which Gopnick and Meltzoff’s proposal needs.

Furthermore, results from the modelling by Plunkett et al. suggest that it may be unnecessary to invoke separate processes or mechanisms to account for the comprehension-production asymmetry. The explanation may lie in the differences between the models of representation in the visual and the auditory/linguistic system. Within the constraints imposed upon the model, it is clear that new behaviours do not necessarily require new mechanisms. Systems integrating information across
modalities can reveal surprising emergent properties that would not have been predicted on the basis of exposure to one modality alone.

**Acquisition of syntax**

How does the young English language learner come to recognise the difference in the message conveyed by the two strings ‘The boy bit the dog’ and ‘The dog bit the boy’ when he/she does not know what a subject, verb or object is? Another difficulty facing the child is that adults often fail to speak in well-formed grammatical sentences. Their linguistic performance is characterised by false starts, slips of the tongue and half-completed sentences, all of which make the grammatical structure of the sentence more difficult for the child to decipher. Acquiring a grammar, given these impoverished learning conditions, would seem an almost impossible task.

Chomsky proposed that the child comes equipped with innate knowledge which directs him/her to look for the special type of rules which underlie the language listened to. This innate knowledge cannot be specific to any particular language – a Japanese child brought up by English speaking parents will learn English just as quickly as an English child will. Chomsky proposed that children come equipped with a ‘Universal Grammar’ – a grammar which underlies all the languages of the world. Knowledge of ‘Universal Grammar’ enables the child to avoid the traps posed by the impoverished stimulus and assists the child in discovering the rules which are particular to his/her native tongue. More specifically, ‘Universal Grammar’ predisposes the child to search for hierarchical structures in his/her language of the type depicted in Figure C.12. These tree diagrams are called ‘phrase structures’.

![Figure C.12](http://scholar.sun.ac.za)

**Figure C.12** A simple ‘phrase structure’ for the sentence ‘The boy kicked the ball’. (NP = noun phrase; VP = verb phrase; Art = article) (McLeod et al., 1998)

The computational machinery which is needed to drive this learning process would seem to rely on built-in parts. It is often assumed that some kind of pre-wired symbol manipulation system is
exploited by the child during the process of language learning. This symbol manipulating system embodies the child’s knowledge of ‘Universal Grammar’. However, the idea that a child might be able to extract grammatical knowledge from the language that he/she hears without prior knowledge of ‘Universal Grammar’ has re-emerged. The perspective has been stimulated by connectionist modelling using simple recurrent networks (SRNs). These take sequences of items and extract the underlying regularities which are inherent in the sequences. Elman (1990, 1993) described a SRN which succeeded in assigning words to grammatical categories (such as noun and verb) on the basis of distributional evidence extracted from strings of words which were based on a set of grammatical rules. In later work, he also demonstrated that recurrent networks are able to learn long-distance grammatical dependencies.

The success of SRNs in correctly predicting long-distance dependencies is determined by the way in which the system is trained. If the network is exposed at the start of training only to simple syntactic forms and then gradually exposed to an increasing proportion of complex structures, final performance on complex forms improves dramatically. The simpler sequences provide the network with an opportunity to discover the basic building blocks of the sentence structure. However, it is by no means clear that children are exposed to simplified input or that such input is beneficial to the child. Therefore, Elman (1993) investigated an alternative procedure which might benefit the network in its efforts at syntactic discovery. Instead of manipulating the training set, Elman manipulated the memory span of the network for previously processed words. Specifically, the activities of the context units were flushed after every third or fourth word by resetting their activities to zero. This removed any memory for words occurring more than three or four items earlier in the sentence and prevented the detection of structure in long sentences. As training proceeded, the memory span of the context units was gradually increased until there was unlimited feedback.

In this manner, Elman was able to demonstrate an advantage of limited memory span in SRNs for the initial extraction of syntactic structure and suggested that ‘starting small’ may be crucial for children’s acquisition of syntax. The complementary nature of the solution highlights the way that nature and nurture can trade off against each other in the search for solutions to complex problems. In one case, environmental factors assisted the network in solving the problem and in the other case, processing factors pointed the way to an answer. In both cases, the solution involved an initial simplification in the service of long-term gain. In development, big does not necessarily mean better.
C.3.3 Cognitive development

Summary

Since the pioneering work of Piaget, one of the ideas which has dominated experiment and theorising in developmental cognitive psychology is that cognition develops in stages. The claim is that as a child matures it acquires qualitatively different processing mechanisms. Thus, the sort of problem it can master will depend on its current stage of development. For example, Piaget believed that an infant’s cognitive processes are initially non-symbolic. In consequence, its mental operations are restricted to processing the stimuli currently impinging on its sensory surfaces. So, in a famous example a child at this first stage of development would be unable to perform a task which required a representation of an object to be maintained after it had disappeared from view. The next stage of development sees the emerging ability to use symbols. The child can now represent stimuli which are no longer present and so can perform tasks, such as responding to an invisible object, which were impossible at the previous stage.

In connectionist models of two tasks, which were originally used by Piaget to illustrate his general theory that cognitive development takes place in stages, the performance of the model undergoes stage-like change, like the child’s. However, the basic architecture of the model remains constant over the course of training and the same learning rule is used.

Development of object permanence

Shortly after birth, infants can track a moving object but they do not appear to be aware that the object continues to exist once it has passed from view. ‘Object permanence’ is the name given to the understanding which the child eventually acquires that an object may continue to exist even when it is no longer visible. Piaget’s views on the development of object permanence came from the study of infants reaching for hidden objects. Initially an infant will reach for a desirable object when it is visible but stop reaching if an occluding screen is lowered in front of the object. By around nine months of age an infant will continue to reach for an object after it has disappeared from view. He concluded that it was not until this age that internal representations of objects were of a form which could be maintained in the absence of an external stimulus.

However, a different experimental paradigm suggests that younger infants already have some form of object permanence. Studies which use the expression of surprise rather than active reaching as a
measure of knowledge, show that infants at around four months understand that an object continues to exist when it disappears from sight (Baillargeon, 1993). The implication is that representations of an object’s properties which could determine its capacity to block the movement of other objects are already in place by four months of age, and that such representations continue in the absence of direct visual input.

Why should an infant cease to reach for a concealed object, as Piaget found, if it has sufficient knowledge of hidden objects to show surprise if their properties are violated, as Baillargeon found? A possible resolution comes from considering differences in the nature of the tasks which infants are required to perform in these two experiments. In Baillargeon’s experiment, the infant is a passive observer, tracking a moving object. The response of surprise indicates that its expectation that the object would reappear at a particular time and place has been violated. In Piaget’s task, the infant is required to make an active response to produce the reappearance of the occluded object. Reaching for an unseen object requires coordination of representations both of where the object is and of its identity (since the infant will only reach for an object he/she wants). Baillargeon’s task, in contrast, only requires the detection that an expectation had not been fulfilled.

Mareschal et al. (1995) constructed a connectionist model to study the development of the representations of object identity and position which would be necessary to perform analogues of Piaget and Baillargeon’s tasks. The model processes the image of a variety of objects as they move across a plane. An object may disappear behind an occluding screen but will eventually reappear on the other side. The model learns to build representations of the objects and their motion which would guide reaching for them (as in Piaget’s task) or express surprise if they failed to reappear from behind an occluding screen (as in Baillargeon’s task).

The model is shown in Figure C.13. Input comes from a retina consisting of a 25×4 array of cells. A typical retinal input sequence is shown in Figure C.14 showing the image of an object covering a block of 2×2 retinal cells moving from right to left at successive times $t_0$ to $t_5$. At $t_1$ it starts to move behind a 4×3 occluding screen and starts to reappear at $t_4$. The trajectory prediction network learns to predict the next position of a moving object. The output from the retina is processed by two separate pathways labelled ‘what’ and ‘where’ in Figure C.13. Neurological evidence shows that in the primate visual pathway, information which is used to establish an object’s identity is processed predominantly along a route from primary visual cortex to the temporal lobe; information about an object’s position and movement is processed predominantly by a route which goes from...
primary visual cortex to the parietal lobe. The model imitates the separate development of representations for object identity and motion.

![Figure C.13 Network used by Mareschal et al. (1995) to simulate the development of object permanence](image)

**Figure C.13** Network used by Mareschal et al. (1995) to simulate the development of object permanence

![Figure C.14 The retinal input and prediction of the position of an object by the network of Mareschal et al. (1995)](image)

**Figure C.14** The retinal input and prediction of the position of an object by the network of Mareschal et al. (1995)

The trajectory prediction module (the ‘where’ pathway) tries to predict the next position of the object. The ‘what’ pathway tries to form a spatially invariant representation of the image so that it can recognise an object irrespective of its position on the retina.

To reach for an object the network would need to integrate information about both the object’s identity (because the child will only reach for an object he/she is interested in) and its position.
The reaching response network is designed to integrate the internal representations generated by the object recognition module and the spatio-temporal representation in the trajectory prediction network. It obtains the spatio-temporal information by sharing the hidden units in the trajectory prediction network. In the model, the task of the reaching network was to output the next position for the objects (the desirable ones) and to output nothing for the other objects (the undesirable ones).

The results from the modelling process mirror the developmental lag observed between Baillargeon’s and Piaget’s experiments. Infants demonstrate knowledge about the properties of occluded objects at a time when they are unable to act towards them. The simulation suggests that the problem is not acquiring the knowledge itself, but the combining of knowledge from two different sources. The network can reach towards a desired object (when it is visible) and has knowledge about the reappearance of occluded objects. What it has difficulty in doing is combining the two sources of information.

This model provides a working implementation of a theory about how infants learn to track and reach for visible and hidden objects. The theory identifies a set of tasks which the model must perform and the information-processing capacities required to perform those tasks. The model is able to make correct predictions about the order of mastery of the different tasks. The fact that the model predicts the order of emergence of various visual motor skills in children offers support for the initial insight that different aspects of children’s object representations develop independently and that this independent development shapes infant performance.

The model can do more than merely mimic previously known facts. It is able to generate new experimental hypotheses which can be tested. The model processes spatio-temporal information independently of feature information: trajectory prediction is done by the ‘where’ channel, and object identification by the ‘what’ channel. The model therefore makes the surprising prediction that in tracking mode there should be no surprise if an object changes its features when it reappears from behind the screen. It should only show surprise if the object reappears at the wrong time. Results from Baillargeon and from Xu and Carey (1996) have shown that infants even as old as 12 months show no surprise at changes in the object when it reappears, provided something appears at the right time.

The critical characteristic of the approach taken by Mareschal et al. was the postulation of different mechanisms tuned to different aspects of the environment. The model assumes the existence of a
mechanism designed to compute object identity and a mechanism designed to track object position. Each network learns independently from the same experience. The asymmetry in performance of the reaching task and the predictive visual tracking task is a direct consequence of the requirement that computations delivered by both mechanisms need to be integrated for the former task but not for the latter. The implications of this conclusion extend beyond the domains of visual tracking and reaching. Any task that requires the integration of the computations from distinct networks is likely to be developmentally delayed compared to tasks which require the computations to be delivered from one network on its own. Given the extent to which the cognitive system appears to have developed along modular lines (Shallice, 1988), many tasks are likely to require integration of information from different sources. This may offer an explanation for developmental lags in other areas of cognitive development.

The separation of the ‘what’ and ‘where’ channels in this model is based on knowledge of the primate visual system. Nevertheless, one can reasonably ask how the networks got to be that way. One answer has been to propose that organisms come equipped with a range of expert networks that possess specialised computational capacities. Tasks are co-opted by the specialist networks depending on the suitability of their computational properties for the task at hand. Jacobs et al. (1991) have shown how mixtures of expert networks exposed to a ‘what/where’ problem of the type described here will always assign the ‘where’ task to the expert network which possesses a linear activation function. The implication is that networks do not necessarily need to be designed to carry out particular tasks. Rather, the task will select the network which has the appropriate (or innate) computational properties.

**Balance beam problem**

A second task used by Piaget to support the claim that cognitive processes develop in stages was the balance beam problem (Inhelder and Piaget, 1958; Siegler, 1981). Children are shown a balance beam with varying weights at varying distances from the fulcrum. They are asked to judge whether it will tilt when the beam is released and if so, which side will go down. McClelland (1989) trained a connectionist network to perform the balance beam task. The contrasting aspects of children’s information processing which Piaget referred to as ‘assimilation’ and ‘accommodation’ can be seen at work in the balance beam model. Assimilation describes the fact that incoming information will be handled by the cognitive structures which are in place at that time and is represented by the consistent pattern of responses of the network within a stage of development. Accommodation describes the fact that existing structures can become modified by new information and is
represented by the gradual changes in the connection matrix of the network which leads to the transition between stages of development. Thus, a connectionist model can provide working examples of concepts which have proved difficult to define precisely without an explicit computational framework.

One of the questions motivating modelling of the balance beam problem was whether stage-like behaviour could emerge from a mechanism which does not undergo any change of learning algorithm or representational mechanism. Connectionist models of cognitive development employ a constant and continuous learning algorithm that computes small changes to the connection strengths in a network to reduce the output error for any given input pattern. Learning in these networks is a process of gradient descent on a multi-dimensional error landscape. If the value of the weight matrix puts the state of the system on the edge of a precipice, a small change in weight values can lead to a dramatic change in the behaviour of the system. Other regions of the state-space will be relatively flat and the behavioural consequences of weight changes will be comparatively minor. Learning in networks can result in periods of stable behaviour interrupted by sudden change, even though the underlying mechanism for learning is one of small continuous change.

Variability in learning

Given a specific start state and training environment, the final state of the networks discussed in this appendix is determined. It would be wrong to conclude that connectionist models necessarily predict a consistent and universal developmental trajectory. It can be demonstrated that:

a) Large individual differences in network learning can result from subtle changes in the training environment and parameters of the learning mechanism;

b) Critical periods in development can emerge as an inherent property of the learning process itself rather than from the onset or offset of some factor that influences learning.

The learning algorithms used in connectionist models include a learning rate parameter. This determines how big the weight changes will be on any learning trial. Varying the learning rate within a simulation (for example, reducing the learning rate as training progresses) approximates the assumption that individual learning experiences are less salient during later learning than early learning. Large learning rates do not always result in overall faster learning. If the learning rate is large, weight adjustments made for one pattern can wipe out much of the learning for patterns processed on previous trials. This is referred to as catastrophic interference.
One might assume that the learning rate in the network has a psychological counterpart in the child that roughly translates into sensitivity to learning experience. Children who are over-sensitive to individual experiences may find mastery of particular types of task more difficult. More generally, these models suggest that substantial individual differences in performance on a task may result from small variations in learning sensitivity.

All the networks described so far have involved static architectures. That is, the numbers of input units, hidden units and output units do not change during the course of learning. However, some types of network models allow new nodes to be recruited during the process of learning. Selective pruning of connections and nodes can also occur in response to their lack of participation in any processing or new learning (Quartz and Sejnowski, 1997). Neural networks which involve node creation and selective pruning may turn out to offer powerful ways to investigate the range of variation apparent in the developmental process.

It is quite common when building a connectionist model to explore the number of hidden units which are necessary to discover the solution to a particular task. It often turns out that there exists an optimum number. One can consider the number of hidden units available to a network as a measure of its representational resources. If the network has inadequate representational resources then it fails to solve the problem. If the network is over-endowed with representational resources then it is likely to get bogged down in memorising individual instances and fail to extract the essential aspects of the problem which would enable it to generalise novel stimuli.

It may be assumed that children vary in the level of resources they invest in a particular problem domain. Connectionist models predict that increased resource allocation does not necessarily lead to a general improvement of performance. For example, over-allocation of resources may assist the child in learning a task but hinder the child in generalising the knowledge obtained by new examples.

Most networks start out life with a well-specified architecture and training environment but a randomised weight matrix. The random configuration of the weight matrix reflects the fact that the network lacks any knowledge about the nature of the problem at the start of training. It is a common experience of network modellers that different randomised start states can have dramatic effects on the learning profile and the final performance of the network. Any configuration of the weight matrix in a network corresponds to a position on the error landscape. Training the network corresponds to traversing the landscape, continually attempting to move downhill to reduce the
error. If lucky, the initial configuration of the weight matrix may be close to a solution to the problem in which case only a limited amount of training is required to achieve a mature state. Alternatively, the initial network configuration may be a long way from a global minimum and extensive training will be required.

It has been shown how connectionist networks might gradually evolve initial weight matrices that are well adapted to learning certain types of task through natural selection from a pool of slightly mutated organisms. The gene pool (in this case the inherited starting weight matrices) will have a degree of variation that will give rise to individual differences in the behaviour of those organisms. Likewise, one may think of variations in the start state of connectionist models of development as an attempt to capture the individual variation of young children in the pre-disposition to learn a task. Connectionist modelers often run multiple replications of a simulation to evaluate the variability of the learning profile and the end state. The variability observed can be used to generate predictions about variability in the modelled population itself.

Critical periods

A common assumption in developmental psychology is that the capacity to learn does not remain constant throughout life. The greater plasticity of the brain for new learning during early development is most apparent in studies which compare recovery from some form of deprivation, trauma or brain damage in early life with that in later life. Similar conclusions are also reached from studies which compare the level of mastery in complex skills such as language when begun early in life with the level of mastery achieved when learning begins later in life. It is often argued that the period prior to puberty constitutes a critical period for language learning, after which time any attempt at complete mastery of a language is unlikely to succeed. Critical periods are often assumed to be domain specific. For example, learning the simple association involved in imprinting may be typical of the very earliest stages of development, whereas the learning of complex cognitive functions may cover a much wider developmental span.

Many explanations of the existence of critical periods in development appeal to maturational factors that impose external restrictions on the ability of the brain to learn. For example, constraints on synaptogenesis, and crystallisation of neural structures as a result of neuronal pruning, provide accounts for the increased difficulty of learning as development proceeds. However, this type of explanation provides little help in understanding why critical periods of development may vary so widely from one domain to the next. In contrast, connectionist approaches to learning provide a
framework for understanding how critical periods in development may result from the learning process itself and are thereby tightly linked to individual tasks. In connectionist models, no extrinsic mechanism is necessarily required to explain why new learning may get harder as development proceeds.

The slope of the activation function plays an important role in determining the size of the weight change that takes place on any learning trial. The slope of the sigmoidal activation function is steepest where the net input is closest to zero, and flattens off when the input becomes large. This means that weight changes are likely to be bigger for units with a net input which is close to zero than units which receive large positive or negative input. Therefore, units with a sigmoidal activation function are maximally sensitive to learning when the input is uncommitted in either the positive or negative direction. These units possess a window of sensitivity for learning. This window is initially open in a network when the weight matrix is randomised at the start of learning. As connections increase in strength, the chance of the net input to a unit being close to zero decreases and the unit moves outside its window of sensitivity to learning. In fact, if the net input becomes very large, then the activity will reach a point where the slope of the activation function approaches zero. Under these circumstances, even a large error signal will produce little change in the connections attached to this unit. The network has already committed itself to a particular representation of the problem and its weight matrix has become fixed in a position from which it cannot easily escape. The network has passed through its critical period of learning.

There are several characteristics of this process which are worth noting. First, it is not all-or-none. The strength of the connections associated with a unit will change gradually during the course of training, the speed of change depending on factors like the learning rate. As long as the unit remains within its dynamic range, the network will remain plastic to learning. Second, for some types of problem, the network never moves out of its window of sensitivity to learning and so never exhibits any critical period effects. This will be true for problems which do not provoke large weight changes – typically, problems which do not contain contradictory examples. Face recognition might constitute an example of a consistent problem domain which normally remains plastic throughout life. In contrast, languages are not consistent in the manner in which they organise the lexicon or exploit grammatical structure. It is to be expected therefore that language learning would exhibit critical period effects. Connectionist modelling offers a natural framework for investigating non-linear patterns of learning where development passes through a critical period as well as learning capacities which remain more stable over the life-span. The power of
connectionist models to deepen understanding on these aspects of the dynamics of development has yet to be fully exploited.

C.3.4 Lesioning connectionist networks

It is easy to simulate the effect of brain damage on a cognitive process with a connectionist network. Once it has learnt how to perform the task, a proportion of the units or connections are removed and the effect on the network’s performance is determined. Lesioned connectionist networks can simulate patient data, both at the qualitative level, reproducing patterns of impairment, and at the quantitative level, predicting the relative magnitude of different impairments. The performance of damaged networks mimics that of brain damaged patients so successfully that this area provides some of the most persuasive evidence that the connectionist approach to understanding cognitive processes is an appropriate one.

One line of evidence in support of the connectionist approach is that, when damaged, the performance of both brains and connectionist networks degrades gracefully. That is, damage typically leads to partial loss of computational power rather than catastrophic failure. Approaches to cognition which do not use parallel distributed processing are usually unable to reproduce graceful degradation. The effect of damage leads to all-or-none functioning which is unlike the graded degradation which is often observed after minor damage to the brain. With a traditional box-and-arrow model of a cognitive process, one can say that a pathway is ‘damaged’ but this is usually little more than hand waving. There is no way of predicting other than in vague qualitative terms (such as the less than illuminating prediction that ‘performance will get worse’) what the effect of a lesion would be. Neuropsychological studies of patients have revealed a wealth of detail about patterns of cognitive impairment following brain damage.

Simulation of deep dyslexia

The breakdown of reading following brain damage, acquired dyslexia, followed many different patterns. An intriguing one is the condition known as deep dyslexia. The distinguishing errors made by these patients are semantic. When asked to read a word aloud, they might say something which is semantically related to the stimulus but has no phonological similarity to the correct response. For example, when shown NIGHT, one patient responded ‘sleep’. The stimulus word must have been processed because the response was semantically related to it. However, all record of its non-semantic attributes has been lost as there is no phonological relationship between
stimulus and response. The traditional interpretation of this would be that there are separate semantic and non-semantic routes from orthography to phonology, and that the patient had lost the non-semantic processing route. However, deep dyslexics also make visual errors, such as responding ‘sandals’ to SCANDAL. In this case, the patient appears to have relatively intact visual information about the stimulus but lost semantic information. Therefore, the traditional interpretation would be that the patient had lost the semantic processing route and kept the non-semantic one. To add to the confusion, deep dyslexics also produce errors such as responding ‘skirt’ to SHIRT. These suggest that they have kept partial information about both the appearance and the meaning of the stimulus rather than losing one and keeping the other. An intriguing error of this sort is one in which there appear to be consecutive visual and semantic errors made to the same stimulus, such as responding ‘orchestra’ to the word SYMPATHY. Presumably, this was a visual error from SYMPATHY → symphony, followed by a semantic error from symphony → ‘orchestra’.

The co-occurrence of apparently independent error types is difficult for traditional box-and-arrow models to explain. The semantic and visual errors suggest the functional separation of sub-systems involved in reading, one semantic and one non-semantic. When a single patient shows evidence of damage to independent parts of the system the modeller is forced to assume that independent parts of the system have been damaged simultaneously. As the number of different error types grows and consequently the number of independent areas which appear to have been damaged, the assumption can start to strain credulity.

A study by Hinton and Shallice (1991) shows how a lesioned connectionist model of reading can start to offer an explanation for the pattern of errors shown in deep dyslexia. The crucial result is that a single lesion anywhere in the model would produce both visual and semantic errors, as well as mixed visual-and-semantic (SHIRT – ‘skirt’) and visual-then-semantic errors (SYMPATHY – ‘orchestra’). This shows that different error types do not necessarily mean that different parts of the underlying system have been independently damaged. A single lesion can lead to different sorts of error.

Hinton and Shallice showed that the co-occurrence of visual, semantic and mixed errors, far from being mysterious, is an inevitable consequence of lesioning a connectionist network which uses attractors to map orthography to semantics. Since this same error pattern is observed in patients, it encourages the view that distributed attractor networks are an appropriate model for human information processing. However, many arbitrary choices were required to make the model run.
Another general problem is that the model was trained with the back-propagation algorithm. This is known to be very powerful but it is not known whether it is used in the brain. Plaut and Shallice (1993) explored whether the result was dependent on these two issues.

All their networks were trained in the same way as the Hinton and Shallice network and then lesioned. The exact pattern of errors varied from network to network, but the crucial result was the same for all of them – all lesions produced visual, semantic and mixed errors. This suggests that it is the existence of attractors (which were formed by all networks), not the details of network architecture, which were essential to the result.

Plaut and Shallice tested whether the use of the back-propagation learning algorithm was crucial by teaching a network with the Contrastive Hebbian learning algorithm and then lesioned it. At a qualitative level it produced the same mix of error types as before, and this mix was achieved wherever the lesion. Thus, the basic result was not determined by the nature of the learning algorithm.

**Double dissociation**

Cognitive psychologists claim to have found a double dissociation when they find one patient who can perform task ‘A’ but not task ‘B’ and a second patient can perform task ‘B’ but not task ‘A’. A natural interpretation of this pattern of deficit is that the cognitive functions which perform the two tasks are independent. The logic and limits of this inference, and examples of its use to support box-and-arrow models of cognitive function are discussed in detail in Shallice (1988). Connectionist simulations of damaged cognitive systems allow an examination of how a particular deficit occurs. They have shown that double dissociations can arise for a variety of reasons other than the existence of independent processing structures. The inference from double dissociation to the existence of independent processing structures underlies many traditional models of cognitive function. The demonstration of double dissociation following selective lesions within a single unified processing system has led to radical rethinking in cognitive neuropsychology.

The double dissociation found between patients in the readability of abstract and concrete words has led some cognitive neuropsychologists, following the traditional interpretation of double dissociation, to claim that there are separate semantic stores for abstract and concrete words. The results from a connectionist model by Plaut and Shallice (1993) show that a single system,
representing both abstract and concrete semantics, can produce either an abstract or a concrete advantage when lesioned in different ways.

A second example of connectionist modelling of cognitive neuropsychological data is a study by Farah and McClelland (1991), who lesioned a model of semantic memory offering an explanation for an apparently anomalous pattern of results found with patients. Separate areas in the brain analyse visual, proprioceptive and auditory input. So one might expect semantic memory to be modality specific, one part storing visual information about objects, one part storing information about actions that were appropriate with them and so on. Support for this view comes from patients who have a deficit related to a specific input modality. For example, visual agnosics cannot recognise an object by sight but can make an action appropriate to the object if they feel it. Apraxics can recognise an object by sight but cannot perform an appropriate action with it.

However, there are patients who show a pattern of deficit which would not be expected were the semantic system modality specific. They show category specific impairments where the information lost cuts across modalities. Warrington and Shallice (1984) reported that it appears that their patient has lost information about living things without suffering any loss to her information about non-living. Her loss appears to be of a category of things rather than of a modality of information. Other patients have been reported who show the opposite pattern – a loss of knowledge about non-living things while knowledge of living things remains intact. The traditional interpretation of this double dissociation would be that there is a functional separation between knowledge about living and non-living objects in semantic memory. This would imply that semantic memory was organised by category. Organisation by modality and by category would seem to be mutually exclusive. Thus, how can evidence from patients be found to support both modes of organisation? Farah and McClelland offered a resolution of this problem by showing that a semantic memory that is organised by modality can, nevertheless, show category specific impairment when lesioned.

A key component of the simulation is the representation of living and non-living items in semantic memory. To try and mimic the coding in people, and hence presumably of the patients who showed the living–non-living dissociation, they asked people to identify any visual or functional descriptors in the dictionary definitions of the living and non-living things used by Warrington and Shallice. Overall they reported about three times as many visual as functional descriptors, so Farah and McClelland divided the 80 semantic units up in the same ratio, 60 visual and 20 functional. They also marked relatively more visual than functional descriptors for both classes, 8:1 for living
and 3:2 for non-living. To mimic these ratios, living things were represented by an average of 16 visual and 2 functional units; non-living by an average of 9 visual and 7 functional units.

A lesion of the functional semantic units in their model affects recognition of non-living things but not of living; a lesion of the visual semantic units produces the opposite pattern. This shows that the double dissociation observed in patients does not force one to the interpretation of independent information storage for living and non-living things. It could arise from damage to a modality specific storage system if the coding for the two categories, living and non-living items, varied in its demands on the different modality specific components. This interpretation is similar to that of the double dissociation found by Plaut and Shallice for the retrieval of concrete and abstract words. If different information processing operations use independent processing systems then selective lesions could lead to the observation of double dissociations in patients. However, different forms of coding or representation can also lead to double dissociations.

Lesioning visual units had a major effect on the recall of information about living things and a minor effect on non-living; lesioning functional units also had a minor effect on the recall of information about non-living things but no effect on knowledge about living things. This is an intriguing result because an asymmetry of just this sort can be found in the patient data. It might seem likely that the asymmetry is the result of differential coding. Living things are represented largely by visual features, so their recall might be expected to be affected only by lesions to visual semantics, while non-living things are represented by both visual and functional semantics so it might be expected that they would be affected by lesions to either part of semantics. A strength of connectionist models is that it is possible to test such assumptions directly.

Modelling an information processing deficit in schizophrenia

Another example of an information-processing deficit is exhibited by schizophrenics. The model by Cohen et al. (1990) makes quantitative predictions of the effect of lesioning the information processing system in different ways and does more than just reproduce the patient data. By comparing the deficits predicted following different lesions with the patient data, the model can be used to decide which component of the information processing system is the most likely source of the deficit in the patients.

The sensory surfaces of humans are continually bombarded by stimuli from different sources. Most of the time, these are of no particular importance and attention can safely wander from source
to source. However, occasionally a stimulus which carries some useful information will appear. An efficient information processing system needs to be able to lock onto a source which is providing important information and prevent distraction from rival sources which are providing uninformative stimuli. The mechanism which does this, is called selective attention. It has been claimed that selective attention is one of the information processing mechanisms which is deficient in schizophrenics.

In one experimental paradigm for studying selective attention, known as Stroop, subjects are presented with words written in coloured ink. They must report the colour of the ink and try to ignore the word. The crucial observation is that people are slower to name the colour of a stimulus like RED than the colour of a stimulus like BED. The responses should be ‘Green’ in both cases if they were written in green. Little interference from the confusing word implies good selective attention and a lot of interference implies poor selective attention. One line of evidence that schizophrenics have problems with selective attention has come from experiments using the Stroop paradigm.

![Figure C.15 A network model which performs the Stroop task (Cohen et al., 1990)](image)

The model of Cohen et al., shown in Figure C.15, reproduces a wide range of effects shown in Stroop experiments with normal subjects. To mimic the performance of the schizophrenic subjects the model was lesioned. They compared the result of lesioning it in two ways: Reducing the rate at which all the units in the network increased their activity in response to an input, or just reducing the input from the Task demand units to the rest of the network. The first mimics a general deficit in information processing, the second a deficit in selective attention. The results from the simulation suggest that it is more plausible to attribute the deficit shown by schizophrenics in the Stroop task to damaged selective attention than to a general deficit in information processing.
C.3.5 Mental representation: rules and symbols in connectionist networks

The traditional approach to modelling cognitive processes is with rules operating on symbols. Since connectionist models have no direct representation of rules or symbols, some theorists argue that connectionist models operate, and thus offer explanations, at a level which is inappropriate for understanding cognition. Some attacks go further and suggest that there are types of rules which cannot be learnt by connectionist models and aspects of symbols which cannot be represented in them. Thus, connectionism is not so much inappropriate, as fundamentally unsuitable. One such problem involves learning to apply a minority default to a novel stimulus. The claim is that connectionist models could never learn to do this because they are constrained to operate by analogy to previous experiences. Minority default mapping can be learnt with an appropriate training environment. The second issue is that many linguistic operations rely on localist representations of concepts. This is easily captured by symbolic representations but seems to be lost in the distributed representations of connectionism. However, symbolic attractors may perform the same function in a connectionist network as the symbols of traditional cognitive psychology.

Learning minority default rules

The natural mechanism for generalisation offered by connectionist models is one of their attractions. However, the bias, which produces similar responses to similar stimuli, can lead to a potentially serious problem. Cognitive mechanisms need to be flexible because domains like language are not consistent. Different responses are sometimes required to similar stimuli and, at other times, similar responses to different stimuli. The tendency of networks to produce similar responses to similar stimuli seems to deny them the required flexibility – hence this problem is known as the ‘tyranny of similarity’.

What would happen in a language in which the default option only applied to a minority of words? The tyranny of similarity would render the connectionist model incapable of producing the default assignment to a novel instance because a novel instance would usually be more like one of the exceptions than one of the words which took the default value. However, minority default assignment is no problem in a rule-based system. The rule produces a response to any word which is not in the list of exceptions. It will work just the same whether 10% of the words are exceptions or 90%. A person learning a language could simply be told the rules that determined which inflection to apply to any verb. Connectionist networks can discover a matrix of weights from an
appropriate set of learning experiences which will produce behaviour equivalent to the implementation of a rule, as demonstrated in models by Hare et al. (1995), and Forrester and Plunkett (1994). Nevertheless, at present they cannot implement the rule just by being told what it is. Since this ability is part of human cognitive skill a full description of the human cognitive system must have some way of doing this.

Symbols and distributed representations

Fundamental to the classical view of cognition is the existence of discrete symbolic entities, representing possible states of the world. On this view, cognitive activity involves the manipulation of these symbols by rules. The interaction of symbols and rules to produce thought can be seen as similar to the interaction between words and syntax which produces language. The problem with connectionism today, on this view, is that distributed representations of propositions lack the internal structure necessary to permit rule-governed transformations. Fodor and Pylyshyn (1988) argue that this counts decisively against connectionist approaches to cognition – aspects of mental processes which they referred to as their systematicity and compositionality.

However, networks with recurrent connections can form basins of attraction so that inputs within a given range will eventually settle on an identical output. Different attractor basins capture different sets of inputs. To the extent that attractor basins are insensitive to small variations in input, they could be considered to have a symbolic quality. Thus, the connectionist equivalent of a symbol is a stable point of attraction in a recurrent network. Rule-governed behaviour might be the trajectory through a series of attractor basins which a network passes through in performing a task such as processing a sentence.

One argument against the connectionist approach is not to deny that cognition is carried out by neural networks in the brain, but to say that studying them is pointless. The networks merely implement rules and symbols in a somewhat opaque way. If one is interested in rule and symbols one should study them directly. A possible counter to this is that although one can study the rules and symbols directly one will never discover anything more than what they are. The connectionist approach may possibly show why they are as they are. For example, a minority default rule could be learnt in a language with one distribution of words and not in a language with a different one.

At the level of rules this would just be a fact; at the level of a connectionist model one can see why. The relative merits of the traditional and connectionist approaches to cognition are affected by what different people see as a plausible scientific theory of the mind.
C.3.6 Network models of brain function

The belief that models of cognitive processes should be brain-like is central to connectionism. Connectionist networks are brain-like at the level of the computational units—they are simple, there are many of them, the way in which the network operates is determined by the strength of the connections between the units, and the network learns and stores information by altering the strengths of the connections. However, in most connectionist models the overall structure has not been related to any particular brain area. Sufficient quantitative information exists about the structure of certain brain areas, and the flow of information to and from them, to build neural network models which have structural as well as neuronal plausibility. Examples include a model of episodic memory formation which has a structure based on the hippocampal system, and a model of visual object recognition which is based on the organisation of information processing in one of the visual pathways in the primate brain. These models involve the learning mechanisms of pattern association, autoassociation, and competitive learning.

Memory formation in the hippocampus

The hippocampus is a structure buried deep in the medial surface of the temporal lobe in the human brain. Its position is shown in the coronal section (or a cut vertically downwards) roughly in line with the ears in the upper part of Figure C.16. The lower part of Figure C.16 shows the structure of the hippocampus in more detail. It consists of two interlocking sheets of cells called the cornu ammonis (CA) (because of its resemblance to a ram’s horn), and the dentate gyrus (DG).

There is a large literature on the memory deficits of patients who have suffered damage to the hippocampus. Although this is complex in detail, some general conclusions are possible. First, the effects of damage to the hippocampus are primarily on the ability to form new memories. A dramatic demonstration of this followed an operation on a patient known in the neuropsychological literature by the initials HM. To control intractable epilepsy he underwent bilateral removal of parts of the temporal lobe including the hippocampus. The operation was successful in controlling the epilepsy but unfortunately led to the patient acquiring acute anterograde amnesia. After the operation he could not form consciously accessible memories of events which occurred in his everyday life. For example, he never recognised the medical staff who looked after him, even though he saw them every day. However, information acquired some time prior to the damage can still be used. Second, within new learning, the effects are selective. Hippocampal damage in
humans leads to failure to form new episodic memories but the formation of procedural memories continues.

Figure C.16 Position and detailed structure of the hippocampus (McLeod et al., 1998)

Episodic memory involves recording of the events which make up day to day experience. It requires the rapid formation of associations between the elements of a particular episode. The result is the formation of a specific memory which can later be recalled by cueing with part of the original memory. A typical episodic memory might involve information about the place where an event occurred, the people who were there and what they said. Cueing someone by reminding them of the place would bring the whole episode back to mind, enabling them to recall the people and what they said. In contrast, procedural memory involves the gradual development over many related experiences of a composite memory from which individual contributing experiences cannot be easily recalled. Procedural memory is typified by skill acquisition. For example, a tennis player gradually develops a memory for the actions required to execute an accurate serve but will not be able to remember most of the thousands of individual practice serves (or specific episodes) that went into acquiring it. The contrast between procedural and episodic memory was exemplified by HM. He could do jigsaw puzzles, and if shown the same one again would do it more quickly, but each time it was shown to him, he would say that he had never seen it before. He could form a procedural memory of the puzzle which could aid in its solution, but he formed no consciously accessible record of the occasions on which he had solved it.
The conclusion from the study of such patients is that the hippocampus is involved in the formation of certain sorts of memories rather than being the actual site of storage in the long term since access to information acquired long before the damage is preserved. The type of memory in which the hippocampus is involved would seem to be that of requiring the combination of information from different sources to form consciously retrievable memories of specific events or facts. There may be temporary storage of these memories in the hippocampus, but since retrograde amnesia is temporally graded, its role as a storage location diminishes with time.

The hippocampus receives input from the parahippocampal gyrus and entorhinal cortex. These areas receive input from virtually all association areas including those in the parietal, temporal and frontal lobes. So the hippocampus has available information from different sensory pathways which has already been cortically processed. There is a divergent set of back-projecting pathways from the hippocampus (via the subiculum and entorhinal cortex) to the cortical areas which provide inputs to the hippocampus, so it would not need to be the long-term depository for the memories which were formed there initially.

It seems unlikely that the hippocampus stores everything that occurs, second by second. If it did, its storage capacity would soon be exhausted, and most of the information it held would be of little interest to the organism. In an effective episodic memory, memory formation would be more likely to occur when something new happened in the environment, or when an event was accompanied by an emotional or motivational response. Events which were accompanied by pleasure, pain or novelty, would be recorded, while the sensory stimulation arising from the repetitive events of daily life would not. The relatively non-specific subcortical inputs to the hippocampus may be involved in threshold setting. They would make it likely that memories would be stored when the organism was in an aroused state, but not when little of interest was occurring. Information about rewards and punishments could be incorporated into the hippocampal memory system via the inputs to the entorhinal cortex from structures such as the amygdala, which are involved in processing information about emotions.

Thus, there is evidence about the function performed by the hippocampus in memory formation, about the cortical areas which project to it and about the neuronal networks within it. Therefore, it is possible to produce a theory of how some memory functions might be performed by the networks found in this brain region.
Input to the hippocampus comes on the perforant path, as shown in Figure C.17. Information processing within the hippocampus occurs in three sequential stages – the dentate gyrus and two areas of the cornu ammonis, known as CA3 and CA1.

![Figure C.17](McLeod et al., 1998)

**Figure C.17** A schematic representation of connections within the hippocampus

![Figure C.18](McLeod et al., 1998)

**Figure C.18** A schematic representation of the flow of information between the hippocampal system and neocortex – thick lines above the cell bodies represent dendrites, and thinner lines with arrow heads the axons. (Δ: pyramidal cell bodies; ○: dentate granule cells; mf: mossy fibre; pp: perforant path; rc: recurrent collaterals of CA3 cells; PHG: parahippocampal gyrus) (McLeod et al., 1998)
Input to the hippocampus from other neocortical areas comes along a group of axons known as the perforant path (1) which synapses with the dendrites of the dentate granule cells and also with the apical dendrites of the CA3 pyramidal cells. The dentate granule cells project via the mossy fibres (2) to the pyramidal cells. The output from the CA3 cells branches – one branch (called recurrent collaterals) forms a set of recurrent connections, synapsing back to the dendrites of other CA3 cells. The other, called the Schaeffer collateral (3), carries the output to the CA1 pyramidal cells, which in turn have connections (4) via the subiculum to other neocortical areas.

The flow of information within the hippocampal system (the entorhinal cortex, areas DG, CA3 and CA1 of the hippocampus, and the subiculum), and between the hippocampal system and the neocortex, is shown schematically on the left of Figure C.18. The solid lines show the flow of information from cortex to the hippocampal system; the dashed lines show the projections back to the cortical areas which originally supplied it with information. On the right of Figure C.18 the projections of typical cells within each area are shown.

The hypothesis is that the perforant path → dentate granule cell system acts as a competitive learning network. Competitive learning removes redundancy, so the output from the DG system will be less correlated and more categorised than the inputs to it from the perforant path. Thus, overlapping signals on the perforant path will be separated before they reach CA3. The role for the DG-mossy fibre system, then, would appear to be to maximize the separation of patterns reaching the CA3 autoassociation system. Recordings from CA3 cells in the primate hippocampus show that the representations of different events do tend to be uncorrelated.

The dentate granule cells which send an input to CA3 produce a sparse representation of the incoming signal to the hippocampus. That is, any given input pattern excites relatively few CA3 cells. In consequence different input patterns are likely to activate different sets of CA3 neurons. A sparse input enables an autoassociator to store more memories. One factor, which limits the number of memories that can be stored in associative networks, is the number of inputs per neuron (Rolls and Treves, 1998). This number cannot be increased beyond about 20 000 in the brain, so the best way to maximize the capacity of the associative memory is to ensure a sparse representation at input. It appears that this strategy is used in the hippocampus.

An episodic memory requires arbitrary sets of concurrent activities to be associated quickly and stored as one event which can be retrieved by a partial cue consisting of a sub-component of the memory. The ability to recall a complex memory with a cue, which is a sub-component of the
whole, is a property of autoassociative memory. The recurrent connections, which are a prominent feature of the CA3 region, mean that this area could act as an autoassociative memory. A new event to be memorised would be represented as a firing pattern of CA3 pyramidal cells. The pattern would be stored using associatively modifiable synapses on the recurrent connections. Subsequently retrieval of a whole representation could be initiated by the activation of some part of it.

**Neural network simulation of hippocampal operation**

The theory of hippocampal function in episodic memory formation was tested by simulating the real network summarised in Figure C.18 with the network model shown in Figure C.19. The aim was to see whether the network could store a large number of unrelated patterns after only a single presentation of each one, and retrieve them from partial clues.

![Figure C.19](http://scholar.sun.ac.za)  
**Figure C.19** A neural network simulation of the hippocampus (McLeod et al., 1998)

To train the model each random binary pattern from a large set was presented once to the entorhinal cortex. A Hebbian learning rule was applied to change the weights of the connections between the input and the DG units which were active. The dentate granule units thus operated as a competitive network, producing different sets of active units for each input from the entorhinal cortex. The output of the DG units activated the CA3 units, which operated as an autoassociation network because of the recurrent collaterals. Hebbian adjustment of synapses took place between the dendrites and the recurrent collaterals. CA3 units changed state as the activity cycled 15 times around the recurrent connections, allowing the autoassociation effect to lead the network towards an
attractor state. The output from CA3 excited the CA1 units and then Hebbian modification of the
connections between the input from CA3 and the CA1 dendrites occurred. The output of CA1
finally provided input back to the entorhinal cortex. Here Hebbian weight adjustment took place
between the output from CA1 and the original input pattern to entorhinal cortex, to implement
pattern association in this pathway. Therefore, the four stages of information processing are
competitive learning in DG, autoassociation in CA3, competitive learning in CA1 and pattern
association between CA1 and entorhinal cortex. Furthermore, units in a network learning by
Hebbian association, and not gradient descent, do not require a continuous activation function.

Parts of each pattern were then presented to the entorhinal cortex as retrieval cues. The ability of
the system to retrieve the whole pattern was assessed by measuring the resultant pattern of firing of
the entorhinal units after the partial cue had activated the network. It is found that recall gets better
from CA3 to CA1 to entorhinal cortex. The reason is that after CA3, there are two sets of
associative synapses, one onto CA1 and another onto entorhinal cortex, and each set can contribute
to improve recall. This illustrates the value of a multistage recall process.

![Figure C.20](http://scholar.sun.ac.za)

**Figure C.20** Performance of a network simulating hippocampal episodic memory retrieval – a
comparison is made for the complete network (entorhinal cortex) and in two earlier
stages, CA3 and CA1 (McLeod et al., 1998)

Figure C.20 shows that a neural network based on the structure of the hippocampus can store a large
number of unrelated patterns with a single presentation of each, and can retrieve the stored patterns
from fragments of them. The immediate interest of a simulation like this is that it works. A
network with the connectivity and general structure of a specific brain area (for example the
hippocampus) presented with inputs which produce a similar sparseness of activity to that found there, using a learning algorithm which is related to one which is known to operate in this area, and given the same limited opportunity to learn a large number of memories (one presentation per pattern) can perform a function similar to that which the area appears to perform in the brain. Surely, progress is being made in understanding how the hippocampus works as part of the brain’s system for memory formation.

However, an implementable quantitative model like this allows more than just showing that a particular approach to implementing brain function works. It allows one to see how performance changes as various aspects of the simulation are changed. Thus, a clearer view of how parts of the system contribute to the performance of the whole can be obtained.

Invariant visual pattern recognition in the inferior temporal cortex

A person can recognise the face of someone known to him/her, whether near or far away, seen full face or at an angle. The pattern of stimulation produced on the retina is quite different in each case. So the visual system must have built a representation of that face which allows recognition to occur independently of the size of the image, its position on the retina or the angle of view. Humans perform visual recognition so effortlessly that this may not seem to be a problem. However, showing how face or object recognition could be performed independent of viewpoint has proved very difficult for cognitive science. To give some feel for the problem it can be shown that the response of a single neuron does not remain constant when a pattern to which it has learnt a response changes position. In consequence, single layer networks cannot perform the apparently trivial task of recognising that two images falling on different parts of the retina are the same. Similar arguments show that neurons do not generalise across scale change. For example, a single layer network cannot tell that the images from an object seen at different distances are of the same object.

However, there are cells in the inferior temporal visual cortex which produce the same response to a particular face irrespective of its orientation or position on the retina. Thus, the problem of computing invariant representations has been solved by the stream of visual processing between the retina and the temporal lobe. The neurophysiology of the stages which visual information goes through from retina to temporal cortex and an implementation of the flow of information through this structure can be summarised as:
a) V1 → inferior temporal cortex is a multi-stage hierarchy. The primary projection for visual information from the retina is to cortical area V1. Subsequent processing of the output from V1 goes on various routes, one of which is via area V2 and V4 to the inferior temporal cortex (IT). Input to a cell in any one stage comes from spatially adjacent areas within the previous stage. Within any layer in the hierarchy there is lateral inhibition between cells.

b) Receptive field size: The receptive fields of neurons in the processing sequence V1 → V2 → V4 → IT become progressively larger at each stage. The receptive field of a visual neuron is the area of the retina in which an event, to which it is responsive, will cause activity in the neuron. Typical receptive fields are around 0.5-1° in V1. At a viewing distance of 1 m a receptive field of 1° covers a region about 1.75 cm in diameter. So cells in V1 are responsive only to stimuli in small regions of the visual field. Receptive field sizes increase to around 8° in V4, 20° in posterior inferior temporal cortex and 50° in the anterior inferior temporal cortex. Therefore, neurons in IT respond to the appropriate stimulus over a large area of the visual field – translation invariance has been achieved.

c) Speed of learning: Learning to identify new objects can occur rapidly. Just a few seconds of seeing a new face or object may enable a person to recognise it later. Although verbal statements about whether an object is recognised may involve processing stages beyond IT, such as the hippocampus, new representations can be built in the inferior temporal cortex in a few seconds of visual experience. This suggests that the learning algorithm should be of the Hebbian type which allows learning to take place in a single trial.

VisNet is a neural network model with a structure based on neurophysiological evidence about information processing in the route from V1 to inferior temporal cortex. The basic structure of VisNet is shown in Figure C.21. Successive layers of 32×32 cells correspond to successive stages in the visual system from V2 to the inferior temporal cortex. (The part of the model corresponding to V1 will be described later). The forward connections to a cell in one layer are derived from a spatially corresponding region of the preceding layer. Each cell in the simulation receives 100 connections from the preceding layer, with a 67% probability that a connection comes from within 4 cells of the distribution centre. The result is that although any cell is only influenced by a relatively small region of the preceding layer, a cell in layer 4 could in principle be influenced by cells anywhere in layer 1. Within each layer there is lateral inhibition between cells. This enables competitive learning to operate locally within each region of a layer. Competitive networks detect correlations between the activity of the input cells, and allocate output neurons to respond to each cluster. These might be thought of as feature analysers.
Each small region of an image presented to VisNet is processed in parallel by a set of filters tuned to a range of spatial frequencies and orientations. The output of these filters forms the input to a competitive net. The activity of these units (layer 1) will represent any correlations discovered among the output of the filters. This becomes the input to another competitive net (layer 2). This will detect correlations among the first set of correlations and, because of the convergent topography in VisNet, operate over a wider area of the original input. This is repeated over four layers. In the final layer, correlations between features in any parts of the retina can be represented. Successive inputs to the network are usually transformations, either rotation or translation, of the same object. The trace learning rule (Hebb-type rule) used, allows correlations represented in the final layer to be between information in images of the same object presented to the network in different spatial positions, or between information about different views of the same object.

Brains, networks and biological plausibility

The two models described use a biologically plausible local learning rule. The signals required to alter the network equivalent of synaptic strength during learning are pre-synaptic and post-synaptic activity. These are both available at the synapse. The systems operate by self-organising competitive learning. This is also biologically plausible as it is driven by the inputs themselves with no external teacher. Competitive learning is implemented by using lateral inhibition which is a well-known property of cortical architecture.

In contrast, many connectionist models do not use local learning rules to alter connection strengths. In back-propagation for example, information about the error in the output units of the network
must be propagated backwards, proportionally to the strength of the intervening connections, and
accumulated at the relevant connection, which may be several layers away, to compute an error
term. Real networks in the brain also differ from many connectionist models in having very large
numbers of neurons, without a bottleneck of hidden neurons, made as few in number as possible, to
ensure that the network learns to generalise well to similar problems. This may be because they do
not have such a powerful learning rule as back-propagation but must instead rely on a local learning
rule. Connectionist models are invaluable in illustrating how phenomena could arise in networks
of simple computing units. Nevertheless, they operate with principles more powerful than those
which are currently thought to be implemented in the brain, so they need to be supplemented by
models based on the types of learning rules believed to be used in the brain, and with the types of
connectivity found in particular brain areas. A challenge for the future is to investigate how the
brain solves difficult problems with less powerful algorithms than those implemented in many
current connectionist networks.

C.3.7 Evolutionary connectionism

All the models described so far start with a clearly defined architecture. The network develops in
response to its environment by adapting its connections following a fixed learning algorithm. The
final state of the network is determined by the initial decisions which the modeller made about its
architecture, the learning algorithm and the environment. However, these do not have to be fixed
at the outset of training. An appropriate architecture and learning algorithm can emerge through a
process of learning, much as the connections in a network adapt gradually in response to the
training environment. Networks can evolve over successive generations to perform better than
their predecessors.

The evolutionary process operates like Darwinian natural selection. A generation of organisms
with some random variation between individuals is exposed to an environment and demonstrate
their relative ability at a task. The more successful organisms are allowed to reproduce.
Reproduction either includes an element of random mutation to ensure variation in the next
generation, or allows mixing of characteristics of successful parents. Successive generations are
exposed to a similar environment and the networks which are chosen to produce the next generation
are selected by the same criterion. As this process is similar to the phylogeny of species, it is
sometimes referred to a genetic connectionism.
Evolution of goal directed behaviour

A hallmark of intelligent behaviour is that it is directed towards a specific purpose – it is intentional. A central problem in understanding the evolution of intelligence is to show how such behaviour might emerge from simpler, non-intentional activities. Nolfi et al. (1994) modelled the evolution of a simple organism’s ability to seek out food. The organism receives sensory input about the distance and direction of the nearest food, and proprioceptive input about its last action. Although the organism has information about the position of food, it has no built-in goal of reaching food and there is no learning algorithm to teach it that goal. Bumping into food is a fortuitous event which will occur if the combination of input signals and weights causes a step in the right direction. Those organisms which collect more food are selected to produce the next generation. The descendants inherit a mutated version of their parent’s weight matrix. It is essential to note that they inherit the initial weight matrix of their parent, not the weights at the end of the exploration period.

The development of goal directed behaviour in these organisms occurs through a process of mutation and natural selection. There is no learning during the lifetime of an individual organism. However, the process of selection and weight mutation across generations achieves the same result as a learning algorithm within an individual organism.

Innately guided learning in speech perception

Newborn infants can recognise speech features within the first few days of life. This has been cited as evidence that the mechanisms for speech recognition must be innate as it seems impossible that learning could take place so quickly. However, there are examples of ultra-rapid learning in animals, such as imprinting and song acquisition. These are usually termed ‘innately guided’. The implication is that the animal has an innate predisposition to acquire this particular pattern of behaviour. It has been proposed that the development of speech perception in infants should be viewed as an innately guided learning process. That is, the system is innately structured so that it is likely to partition continua which form the basis for classification of speech, such as voice onset time, into appropriate categories. Thus, learning the sound properties of the native language can take place very quickly.

Nakisa and Plunkett (1998) developed a genetic connectionist model to investigate what the innately guided and the learnt components of the development of speech perception might be. The
architecture and learning rules of the network were encoded in a 'genome'. New networks were generated by reproduction (random recombination of parts of two 'parental' genomes). The network took speech as its input and variants which succeeded in detecting contrastive features at output were allowed to reproduce. The genetic recombination algorithm succeeded in discovering a network which could detect speech features after only two minutes exposure to spoken English. This remarkable achievement can be viewed in two ways. It gives an insight into the performance of human infants, showing that, the ability to make speech discriminations in the first few days of life does not necessarily imply an innate representation of speech sounds, but with the right architecture and learning rule, very rapid learning is possible. It also demonstrates how useful genetic algorithms can be to modellers. Nakisa and Plunkett started with no clear idea of what would be required for a network to achieve ultra-rapid learning of categories suitable for portioning speech. Rather than trying endless combinations out themselves, they allowed the genetic recombination algorithm to do the work.

The network was trained on a variety of Voice of America broadcast languages – English, Cantonese, Swahili, Farsi, Czech, Hindi, Hungarian, Korean, Polish, Russian, Slovak, Spanish, Ukrainian and Urdu. All of the languages tested were equally effective in training the network to represent English speech sounds. That is, the same initial feature detectors develop whatever language the network is first exposed to. This echoes the observation that infants are initially able to discriminate speech contrasts from any language. For example, 6-8 month old infants from an English-speaking background can distinguish the glottalised velar-uvular stop contrast /ki/-/qi/ in ‘Nthlakapmx’ and the Hindi voiceless aspirated versus breathy voiced contrast /h/-/d/. However, this ability is lost after about a year’s exposure to any language which does not use these contrasts.

Nativism and constructivism

By developing appropriate architectures, time-constants and learning rules over many generations, a system evolves in which the task of learning to represent speech sounds can be achieved after very brief exposure to speech. Nature finds a structure which can learn quickly, but having the correct innate architecture and learning rules is not sufficient for creating good representations. Weights cannot be inherited between generations so the network is dependent on the environment for learning a set of weights which can perform the task. If trained on acoustically filtered (or unnaturally distorted) speech, the model does not form good representations. However, given the sort of auditory input heard by an infant the model rapidly creates the same set of features whatever the language it hears. This demonstrates an advantage of innately guided learning over
conventional self-organising networks. It is less dependent on the ‘correct’ environmental statistics. The model found by the evolutionary process in this simulation offers an account of how infants from different linguistic environments can come up with the same featural representation so soon after birth. In this sense, innately guided learning as implemented in this model is halfway between nativism and constructivism. It shows how genes and environment interact to ensure the rapid development of featural representation of speech on which further linguistic development depends.