An alternative to the Lewisian view of conventions

JP Smit

Department of Philosophy, Stellenbosch University, South Africa
E-mail: jps@sun.ac.za

Abstract

Lewis (1969) characterises conventions as regularities that arise from recurrent coordination games. I argue, contra Lewis, that conventions are rules that promote a relevant goal in virtue of coordinating our behaviour. I demonstrate the virtues of this view by showing that it provides an elegant way of dealing with four basic objections to Lewis's view, namely that Lewis requires agents to understand their own situation too well, that his view robs conventions of explanatory force, that it mischaracterises cases where someone has non-prudential reasons to follow a convention, and that it mischaracterises situations where the relevant behaviour is non-uniform.

Key words: conventions, Lewis, rules

1. Introduction

In *Convention* (1969), Lewis sets out to investigate the platitude that language is conventional. His key idea is that conventions are regularities that result from recurrent coordination games and where the parties to the coordination game are aware of the status of the resultant regularity. In this paper, I claim, contra Lewis, that conventions are not regularities of a certain type, but rules of a certain type. A convention exists in a society when the members of a society are disposed to follow such rules. Furthermore, I will claim that such rules need not resolve coordination games. Rather, it just needs to be the case that such a rule promotes a relevant goal in virtue of coordinating our behaviour. I will defend my proposal by working through four objections to Lewis's view that motivate departing from his views in the way that I propose.

2. Lewis's view of conventions and the nature of the present inquiry.

Lewis analyses conventions as follows:

> A regularity *R* in the behaviour of members of a population *P* when they are agents in a recurrent situation *S* is a *convention* if and only if it is true that, and it is common knowledge in *P* that, in any instance of *S* among members of *P*,
>
> (1) everyone conforms to *R*;
> (2) everyone expects everyone else to conform to *R*;
> (3) everyone has approximately the same preferences regarding all possible combinations of actions;

(4)   everyone prefers that everyone conform to *R*, on condition that at least all but one conform to *R*;

(5)   everyone would prefer that everyone conform to *R'*, on condition that at least all but one conform to *R'*,

where *R'* is some possible regularity in the behaviour of members of *P* in *S*, such that no one in any instance of *S* among members of *P* could conform both to *R'* and to *R* (Lewis 1969: 76).

Lewis (1969: 78) develops a definition later on that allows for exceptions to the strict conditions given above. I will, as most of those writing about Lewis do, focus mainly on the exceptionless version. The core of Lewis' theory is that conventions are regularities that arise in response to game-theoretical coordination problems where participants are aware of the status of the resultant regularity. The defining characteristic of a 'game' as such is that the context of interaction must be *strategic*, i.e. optimal strategy for one party must depend on the behaviour of the other party (or parties) involved. A game is a game of coordination if the interests of the actors are *aligned*, i.e. actors have the same ordinal ranking of the different outcomes[1]. However, Lewis (1969) does not require that the parties to a convention have perfectly aligned preference. Rather, in condition (3), he merely requires that preferences must be 'approximately' the same (Lewis 1969: 76). This allows Lewis (1969: 14) to treat games like 'battle of the sexes'[2] as potentially giving rise to conventions, even though the preferences of the parties involved give rise to differing ordinal rankings of the possible outcomes.

The basic constraint on the game theoretical structure of the interaction that Lewis (1969) imposes depends on his ingenious notion of a *coordination* equilibrium. The standard notion of an equilibrium in game theory is that of a set of strategies such that no actor can make himself better off by unilaterally changing his strategy. Such strategies are said to be in equilibrium as, if the actors somehow hit on such a combination, the outcome thereby reached is likely to be stable. Lewis's (1969) notion of a coordination equilibrium differs from that of a standard equilibrium in that, in a coordination equilibrium, no actor can be made better off by *any* actor changing their strategy. Mutual defection in a prisoner's dilemma[3], for instance, is an equilibrium, but not a coordination equilibrium, as one party can make the other better off by cooperating. Using this notion of a coordination equilibrium, Lewis (1969: 16) ultimately imposes the requirement that the kind of coordination game needed to give rise to a convention must be a game with multiple coordination equilibria.

One way of understanding the conceptual question as to the nature of a convention is to understand it as an analysis of everything to which we apply the natural language term 'convention'. The basic goal of such an enquiry would be to try to arrive at an analysis that

---

[1] Lewis (1969: 13–14) follows Schelling (1960) in distinguishing between games of pure coordination and games of pure conflict.

[2] In a typical 'battle of the sexes', two parties have to choose between going out and staying in. Both would prefer the outcomes where they choose the same option to those where they choose different outcomes. However, one party would prefer both parties going out to both parties staying in, and the other party would prefer both parties staying in to both going out.

[3] In a typical two-person prisoner's dilemma, parties have to choose between 'defecting' and 'cooperating'. Both would prefer to be the sole defector, whereas both need to avoid being the sole cooperator. It is further stipulated that both prefer mutual cooperation to mutual defection. The sole equilibrium (i.e. state with no incentive in favour of unilateral deviation) in a one-shot prisoner's dilemma is mutual defection.

gives necessary and sufficient conditions that apply to everything that a competent speaker of English would call a convention. However, this is not what I will be trying to do. It is far from clear that the things we call a 'convention' form a kind. The things that we most commonly think of when we think of conventions include the matter of driving on the same side of the road in a given country and linguistic rules like referring to Kripke as 'Kripke'. My analysis will apply well to such cases, but there are also other forms of behaviour that we call conventions that it does not straightforwardly apply to. Many will claim that, 'in some sense', fashion is a matter of convention, that working until five o' clock is a matter of convention, that proper table manners and other matters of etiquette are conventional, and so on. While, for instance, how we use language and how we choose to dress are presumably related in some non-trivial way, it seems *prima facie* unlikely that we are here dealing with exactly the same kind of behaviour. Lewis (1969), similarly, was not trying to give an account that would capture every single use of the word 'convention'. Where his theory clashed with ordinary usage he was untroubled, admitting the existence of "genuine usages that do not fall under my analysis"[4] (Lewis 1976: 113). My goal here is similar. I will not be trying to give a theory of everything that can be called a 'convention', i.e. trying to do the job of a lexicographer.

Lewis wrote *Convention* with the aim of arriving at a theory of conventions that can be useful in the study of language. His theory aims primarily at, and is tailor-made for, capturing what we may call 'the conventionality of language'. I will proceed similarly. I will attempt to develop a notion of convention that is important, with this importance being a matter of being useful in explaining our use of language. Hence, in the first instance, the task is not one of analysing what we *mean* by 'conventions', but developing a notion that captures at least some of what makes an action one that accords with what we call a 'convention', and is useful in explaining linguistic behaviour. This task can be glossed as being a matter of 'explaining the sense in which language is conventional'. For this reason, I will continue talking of giving a theory of 'convention', where this task is understood as explained here. I take it that, as explained above, this is what Lewis (1969) was also trying to do.

The difference between what I will try to do and an analysis that captures how we use the term 'convention' should not be overstated. Where possible I will try to develop a theory that is consistent with how we use the term 'convention'. The only difference is that, if our usage differs from the theory to be developed here, but the way in which usage differs has little or no explanatory value when it comes to language or introduces needless complication, I will ignore common usage. Where such considerations do not apply; however, I will try to make the theory consistent with common usage. In fact, I am of the opinion that the view to be defended here sticks closer to how we apply the term 'convention' than Lewis's (1969) view does.

3. Conventions are Coordinating rules

The task to be carried out then, is this: develop a theory of 'convention' that explains our linguistic behaviour and departs from common usage of the term 'convention' only when necessary. In order to state my view in an intuitive form, several notions need to be defined, the first being the notion of a 'rule'.

---

[4] Lewis (1976: 113) suggested that such usage may be derivative in some way, i.e. to depend on his notion of convention in some deep sense. I suspect the same of the view developed here.

> Rule: A rule *R* is an injunction to perform an action *K* if specified circumstances *C* obtain. Rules can be stated in the form 'if *C* obtains, perform *K*'.

Rules are employed by agents to promote various goals. In this way, the rules of valid inference promote truth-preservation, the codified rules of food labelling promote consumer safety, some self-imposed rules of conduct promote productivity, and so on. For current purposes, a goal of action must be distinguished from a mere benefit that is realised in virtue of action. Take for, instance, someone who takes up jogging to improve his health. Further stipulate that the person enjoys meeting the kind of people one meets through jogging, but this consequence of jogging would not have been sufficient to motivate him to take up jogging. Ordinarily, we would not object if the person said that he jogs to improve his health and meet people. On the definition used here; however, only the health benefits of jogging count as a goal of action, the latter is a mere benefit.

We can allow for cases of self-deception or a lack of self-knowledge by not requiring that the goal the person thinks motivates his behaviour, actually be the goal that motivates his behaviour, or that the person is aware of the goal that motivates his behaviour. Consider someone who drives on the left-hand side of the road, thinks that he does so due to religious conviction, and does not realise that, if driving on the left did not enable him to avoid head-on collisions, he would switch to driving on the right. Such a person essentially has an incorrect theory about why he persists in a specific course of action. In such a case the goal of obeying a divine injunction does not, for our purposes, count as a goal of action, while the goal of avoiding head-on collisions does. On the final analysis then, we can define a goal[5] of action as a reason for action that explains why an action is committed, independently of whether the person is aware of the fact that the outcome motivates his action.

All rules promote a goal through a certain *mechanism*, i.e. in a certain way. For example, the way in which the rule 'if at the office, don't use the internet' promotes productivity is that it eliminates one source of distraction, the way in which the rule 'if you experience severe pain while exercising, stop' promotes being healthy is that it stops those who follow it from exacerbating a serious injury, and so on. One mechanism whereby a rule can promote a goal is coordination. Intuitively, a rule that promotes a goal via coordination does so in virtue of making it come about that our actions are similar or differ in some relevant way. More precisely, we can define the notion of a *Coordinating* rule as follows:

> A rule *R* that promotes a goal of action *G* is a Coordinating rule if, and only if,
> (1) *R* is followed in order to promote *G*;
> (2) the effectiveness in promoting *G* of an action that exhibits *R* in a strategic *context* of interaction *C* primarily depends on the number of actions in *C* that exhibit *R*;
> (3) the effectiveness in promoting *G* of an action that exhibits *R* in *C* increases as a function of the number of actions in *C* that exhibit *R*;

---

[5] For an alternative view that defines conventions in terms of (entities like) rules and goals, as opposed to regularities and preferences, see Miller (1992). Miller (1992: 436–437) defines conventions in terms of 'procedures' and 'collective ends'. Though the substance of my theory differs significantly from his, my choice of terminology is not supposed to be indicative of any deep difference on the nature or role of goals and rules.

(4) if *all* actions in *C* exhibit *R*, then there is no action in *C* such that, if replaced by an action that violates *R*, the replacement action would have been more effective in promoting *G*; and

(5) if all actions in *C* exhibit *R*, then there is no action in *C* that exhibits *R* such that it would have been more effective in promoting *G* if some combination of the other actions in *C* violated *R*.

For present purposes, I define the notion of behaviour 'exhibiting a rule' as behaviour that accords with what the rule prescribes, independently of whether we would class the behaviour as rule-following or not. The notion of a strategic context, in turn, is defined as a situation in which the optimal action to perform depends on what other actions will be performed. Note that condition (4) is an adaptation of the standard idea of an equilibrium, whereas condition (5) is an adaptation of Lewis' (1969) idea of a coordination equilibrium. Including condition (4) allows us to exclude rules that advise us to cooperate in prisoner's dilemmas. This is required as such rules are not conventions, but moral norms. Including condition (5) allows us to exclude rules that advise us to defect in prisoner's dilemmas. This is required as, at least in one-off cases, defection is optimal independently of how others behave, whereas, in the case of conventions, the best way to promote a relevant goal is conditional on how others behave.

Note that the above definition of a coordinating rule is not stated in terms of agents who perform actions, but instead in terms of the actions themselves. This is done as, strictly speaking, conventions only require a multiplicity of actions that are strategically related, not a multiplicity of agents. Hence, we should include cases in which the interactive context is an intertemporal one where different actions of the same individual promote some goal in virtue of all such actions being actions that exhibit the same rule. Consider a being who must eat once a day and who can minimise his chances of falling ill by spacing these meals as far apart as possible. This implies that the being should eat at the same time every day, but it does not matter when he does so. If such a being adopts a rule 'every day, eat at noon', then this counts as a coordinating rule, even if only one person[6] is involved. Note that, by the same standard, a secret script that an individual invents in order to keep his diary entries private also counts as a set of coordinating rules.

Most relevant contexts of interaction will feature different agents with each performing an action, and, as formulating such cases in terms of the actions themselves can be somewhat inelegant, I will mostly formulate my claims in terms of agents performing actions. So, turning to some everyday examples, the rule 'if in the UK, drive on the left' is a Coordinating rule on the above definition as: (i) people do so in order to avoid head-on collisions; (ii) the effectiveness of driving on the left in avoiding head-on collisions primarily depends on how many people also drive on the left when I encounter them; (iii) the effectiveness of driving on the left in avoiding head-on collisions increases as a function of how many people also drive on the left when I encounter them; (iv) no person can improve his chances of avoiding a head-on collision by driving on the right; and (v) no person, or group of persons, can improve the chances of anyone who drives on the left to avoid a head-on collision by driving on the right. Equivalent claims are true for the way in which linguistic conventions promote communication, the way in which adopting a given currency lowers transaction costs, the way in which having the initial caller call back when a call is dropped aids speedy resumption of the call, and so on.

---

[6] Alternatively, we could define conventions as holding between 'agents', where agents are objects that can be persons or time-slices of persons.

Note that the above construal of conventions as Coordinating rules is not incompatible with the truism that conventions often advise us to do different things. Even in such cases we all still 'do the same thing' in the sense of all *exhibiting the same rule*. Conventions that advise us to all do different things can be phrased as rules that all can follow, so that it is natural to say we all do the same thing. Even when we say that everyone in the UK 'drive on the same side', this only makes sense if we implicitly interpret 'same side' in terms of egocentric coordinates like left and right. The whole point of such a convention is to ensure that vehicles going in opposite directions drive on *different* sides of the road. If everyone really only ever used the same side, where 'same' is defined without reference to egocentric coordinates, it would have disastrous consequences.

The view I defend is that conventions are coordinating rules. If this is accurate, then the conditions under which a convention can be said to exist are the conditions under which we can say that someone follows a coordinating rule. To do this, we need to first pay attention to the conditions under which someone who should count as a rule-follower of the required type would have a good reason not to follow the rule. We need to distinguish between two kinds of defeasibility, call the first 'internal defeasibility'.

> A rule R is internally defeasible if, and only if, R is followed in order to promote a goal G and there can be occasions where violating R is more effective than following R in promoting G.

All rules are not, of course, internally defeasible. The rules of valid inference cannot be overruled in the required manner by some feature of a specific context. However, a Coordinating rule must always be internally defeasible. Consider the matter of driving on the left-hand side of the road in the UK. As people generally obey it, it is rational to adopt the rule 'if driving in the UK, stick to the left-hand side'. However, there can be occurrences of driving in the UK in which this is no longer an optimal way of not-crashing. I can always encounter some other driver who, either by mistake or not, drives on the right and effectively forces me to break the rule. This is true of all Coordinating rules. The efficacy of Coordinating rules depends on other people also obeying them. This means that there can always be a scenario where others' breaking the rule creates a situation where I have a *prima facie* reason to break the rule. Hence, all Coordinating rules are internally defeasible rules.

Coordinating rules are also defeasible in a broader sense, which I will call 'external defeasibility'.

> A rule R is externally defeasible if, and only if, it is followed in order to promote a goal G and there can be occasions where some goal G' is more motivating than G and achieving G' necessitates violating R.

Consider cases where the specific circumstances force my hand in some way by, for instance, providing a non-strategic reason for action. If, for example, I am driving in the UK and there is no car within a mile from me, but there is a giant pothole in front of me, I have a reason to drive on the right in order to get around the pothole. In such a case, the goal of not-crashing does not guide my behaviour, as the goal of avoiding the pothole is more motivating. The motivating power of a Coordinating rule is always externally defeasible, as the goal of following the rule

can always be superseded by some more motivating goal. Hence, Coordinating rules are both internally and externally defeasible.

With the above notions defined and explained, a relatively simple statement of the conditions under which a convention exists can be given.

> A rule *R* is a *convention* among a sub-group *S* of a population *P*, if, and only if, *R* is a coordinating rule that, absent external or internal defeaters, and absent relevant false beliefs, all members of *S* are disposed to follow.

Call the above view the *Coordinating rule view of conventions*. Some of the reasons for adopting it should already be clear from the above discussion. The clause about the absence of relevant false beliefs is included in order to deal with cases where people try to coordinate their behaviour, but happen to be mistaken in some relevant way. In other words, cases where some person is disposed to drive on the right in the UK in virtue of thinking that driving on the right is the current, dominant driving standard in the UK, or a case where someone uses 'Lucas' to speaker refer to Krugman in virtue of thinking that it is standard usage, etc. In such case, even though the behaviour of the person will not generate the regularity that allows him to be a part to a 'convention', as Lewis (1969) defines it, there is a clear sense in which the person is a party to the convention, despite merely violating it by mistake.

Below I will discuss the objections to Lewis' (1969) account that motivate rejecting his view in favour of the Coordinating rule view.

4.   Objections to Lewis

4.1 Objection 1: The knowledge requirement.

Lewis (1969), in his characterisation of conventions, requires that his conditions (1) – (5) must be common knowledge among the parties to the convention. This requirement is convincingly criticised in Burge (1975)[7]. Burge (1975: 250) points out that we can imagine speakers who are only aware of the existence of one language and believe that the words in the language are somehow 'naturally', or by supernatural fiat, connected to what they mean. Such speakers would not understand their own language use as conditional on how others use language. Yet we would not hesitate to call their language use conventional, despite the fact that that they will explicitly deny its conventional nature.

Burge (1975: 250–251) also points out a deeper problem. Throughout the history of philosophy, many have claimed that certain values and beliefs, thought to be somehow natural, are actually based, in some deep sense, on conventions. In this way, certain basic doctrines in mathematics, logic and ontology have been claimed to reflect human conventions, as opposed to how things objectively are. On Lewis' (1969) construal, such a move would seem to be inherently absurd, as it would be constitutive of conventions that those who use them understand their continued use to be conditional on other people also conforming to the same convention. The claim that such an argumentative move is intrinsically incoherent is implausible.

---

[7] Burge's (1975) argument has generally been found to be persuasive. See, for example, the discussion in Blackburn (1984: 120–122).

The Coordinating rule view does not require the parties of a convention to understand the state in which they find themselves. As such it does not, by definitional *fiat*, rule out the possibility of discovering that some aspect of our behaviour is conventional. Note that this is not to deny that common knowledge has a fundamental role to play in explaining the origin or persistence of conventions, as clearly it does. All that is denied is that the parties to a convention need to understand why they act as they do.

4.2 Objection 2: Rules instead of regularities

Below I will argue that conventions are not regularities, but rules. Before we get to the meat of the argument, note one initial point in favour of such a claim. Conventions, as a quick google search will confirm, are commonly said to be the kinds of things we can follow or violate. If conventions are rules of a certain kind, then this matter of common usage is explained as rules are also commonly said to be the kinds of things we can follow or violate. However, regularities are not the kinds of things that are commonly said to be followed or violated. The expressions 'follow a regularity' and 'violate a regularity' are simply not standard in English. We can, of course, speak of 'actions in accord with a regularity', as we can speak of 'actions in accord with a convention'. But, we can equally well speak of 'actions in accord with a rule', and hence this latter usage does not favour the regularity-view over the rule-view. The basic point is that we typically portray conventions as things that can be followed or violated. This matter of usage is explained by rule-view of conventions, but not on a regularity-view of conventions. This fact should serve to give the rule-view some initial plausibility.

The main aim in this paper, however, is not to explain maters of usage, but to develop a notion of convention that is of use in explaining phenomena like language. To this end, consider the regularity that is supposed to be constitutive of the existence of a convention. In the case of conventions concerning driving, this is a matter of the side of the road that different people choose to drive on being highly correlated. How do we explain this regularity, i.e. the fact that people generally drive on the left-hand side of the road in the UK, etc.? The intuitively appealing answer is that the regularity is explained by the convention of driving on the left-hand side of the road. This would both fit our common usage of the term 'convention' and allow an analysis that uses conventions to have explanatory force. However, Lewis (1969) cannot give this type of answer. On Lewis' (1969) view the existence of the regularity is *constitutive* of the existence of the convention, and hence cannot explain the existence of the convention. Conventions can only explain regularities if they somehow give rise to them. Such an explanation presupposes that the notion of a convention is independent of that of a regularity, and the existence of the convention as prior to the existence of the regularity.

If we wish to save the idea that conventions explain regularities in action, conventions cannot be equated with regularities. Is there a way of defining the notion of a convention so that conventions can have such explanatory force? One way of doing so would be to restrict the analysis to conventions that exist only in virtue of explicit agreements, or promises to act in a certain way. Take, for instance, a case where all drivers explicitly agree to drive only on the left-hand side of the road. We can now simply define the notion of a convention in terms of an agreement to follow the agreed *rule*. This agreement to follow a certain rule then gives rise to the regularity, and hence we save the idea that conventions explain the later regularity in action.

However, the above proposal has obvious drawbacks. It can only account for the special case where explicit agreement gives rise to the regularity. Moreover, the whole point of Lewis' analysis was to show that conventions need not be based on explicit agreements. Fortunately, we can gain the relevant explanatory power of rules without requiring such rules to be followed in virtue of explicit agreements. When people agree to follow a convention, this amounts to an agreement to, when a certain set of circumstances occur, act in a certain way. We can distinguish between different ways that it can come about that people follow a rule. One way of making this come about is by explicit agreement, but this is not the only way that it can happen. Rule-following can emerge spontaneously in any number of ways, most prominently as a response to a coordination problem. This means that we can identify a convention with a certain type of rule, independently of how it came about that the rule is to be followed.

Note that defining the existence of a convention in terms of a *disposition* to follow a rule also secures the result that a convention can exist even if it has not yet been followed. This is a virtue as, if this were not so, it would *never* make sense to explain the first instance of some action in accord with a convention as being due to the convention in question. Consider, again, the case of two people explicitly agreeing to drive on a certain side of the road. Their agreement to drive on the left uncontroversially constitutes a convention. The first time that either of them drives on a road, this act is explained by the existence of the convention explicitly agreed to. But, once again, this can only be so if the convention existed prior to the act of driving. Hence, it cannot be constitutive of a convention that it has been followed, but only that the parties to the convention are disposed to follow it.

Note that defining the existence of a convention in terms of a disposition to follow a rule is useful in explaining, for instance, how a baptism can make it come about that a name conventionally refers to a particular individual. In a typical baptism, a name is mentioned, not used, and so we cannot say, after a baptism, that the relevant convention has been followed. Yet a baptism can make it come about that a person has a certain name, prior to the name actually being used. This is explained by the fact that a baptism can *dispose* people to use a name in a certain way in virtue of making the relevant rule salient. Such a disposition, then, is the fact that is constitutive of the particular individual having the relevant name.

Lewis (1969: 100–107) considers the possibility of defining conventions as rules. Lewis (1969: 104) acknowledges that it is hard to "argue that some conventions are not naturally called rules", but rejects any attempt at characterising conventions as rules. His objection is "that the class of rules is a miscellany, with many debatable members" (Lewis 1969: 105). This is defended by pointing out the many kinds of things that we call 'rules' that are obviously not conventions.

It is hard to see exactly what Lewis's (1969) argument is supposed to be. Lewis (1969) argues by constructing a list of things we call 'rules' and pointing out that most of them are not convention. However, this is a weak objection to a view taking the approach of the Coordinating rule view of conventions. On the Coordinating rule view it is not the case that all rules are supposed to be conventional, just that some are. One could construct an equally weak argument against the regularity-view of conventions by pointing out that there are all kinds of regularities that are not conventions. This argument would be weak, as Lewis (1969) is not claiming that all regularities are conventions, but only that regularities of a certain type are conventions. The Coordinating rule view similarly only claims that rules of a certain type are conventions.

Lewis (1969: 105) also argues that the notion of a rule is "an especially messy cluster concept". This may well be true, but I do not see this as a major objection to characterising conventions as rules. The notion of a rule has been defined here quite precisely as an instruction of the form 'if *C* obtains, perform *K*', where *C* is a situation or context and *K* an action. It takes only a moment's reflection to see that most of what we sometimes call 'rules' (moral norms, rules of etiquette, rules of inference, linguistic conventions, etc.) can be stated in this form, even if the formulation is sometimes a bit inelegant. I do think that uses that do not have this form (e.g. Lewis's (1969: 100) example of it being a rule that all meat is more tender if cooked at low temperatures) tend to be examples of loose usage or to be derivative of the notion I have defined here. But, be that as it may, I will not argue that here, as nothing depends on it. How we use the term 'rule' is, ultimately, a matter of mere lexicography, what matters at present is what conventions are. If the reader is unconvinced that my characterisation of rules reflects common usage, he can simply interpret my use of 'rule' as a technical term that, by stipulation, has the form 'if *C* obtains, perform *K*'. Note that nothing of consequence would change if I were to call the view defended here the 'Coordinating instruction view', 'Coordinating injunction view' or 'Coordinating imperative view', or even make up a new term altogether. Whether conventions have the form 'if *C* obtains, perform *K*' is an important matter of substance, whether we should call anything with such a form a 'rule' is, ultimately, trivial.

4.3 Objection 3: Promises, overdetermination and coordination games

Lewis's (1969: 73) construal of conventions requires, at least in the case of games with discrete moves that the context of interaction amounts to a coordination game in the sense of having at least two coordination equilibria[8]. This requirement leads to a problem when people promise to follow a course of action that we would generally consider conventional.

We can distinguish two cases. Consider, firstly, a society where everyone makes a binding promise, if they drive a car, to do so on the left-hand side of the road. Stipulate that avoiding head-on collisions is reason for action among them, i.e. if they encounter a society that drives on the right, they will keep their promise, not by driving on the left and crashing, but by simply not driving at all. In such a case, Lewis (1969) will not consider their action of driving on the left a convention, as their preference for driving on the left is not conditional and they have no inclination to drive on the right, even if everyone else did so. On the Coordination rule view of conventions this would still count as a convention, as it is still the case that driving on the left satisfies conditions (1) to (3). On the Coordination view, we should simply say that the people involved follow the convention of driving on the left as they had promised.

Consider a second case, where the members of a society promise to drive on the left, meaning that they will do so even if it would lead to crashing their cars. In such a case, Lewis (1969) would not classify their behaviour as conventional, as their preferences are not conditional on how others behave. Here the coordination view would agree that the behaviour is not conventional, but would ascribe this to the fact that avoiding head-on collisions is not a goal of action. Even if the people would, all else being equal, prefer not crashing to crashing, this does

---

[8] Lewis explicitly states that, if the context is not of this kind, then there can no longer be a convention. See, for instance, his discussion of notations (Lewis 1969: 103–104). Here Lewis (1969: 103) states that if a standard notation in logic is enforced by editors with a threat of non-publication, such a notation is no longer a convention. This is both implausible and cuts reality up in a way that makes things needlessly complicated.

not guide their action, and where not-crashing is achieved it would be a mere benefit of action, and not a goal of action. On the coordination rule view, accidentally generated benefits are not the kinds of outcomes that are constitutive of the existence of a convention.

Lewis would see neither case as conventional, while the coordination view would treat the first as conventional, the second not. Note that the first case is essentially a problem of overdetermination. The people on the island have prudential reasons to drive on one side of the island, but also have non-prudential reasons to do so. Promises are not the only complicating factor that raises difficulties related to overdetermination for Lewis' account. Consider the Burge (1975) objection to the common knowledge requirement discussed above, but add the stipulation that we are dealing with a society that believes the rules of language were laid down by God. Furthermore, stipulate that they believe that breaking such rules is punishable by eternal damnation. Once again, in such a case, Lewis (1969) cannot characterise their linguistic usage as conventional, as, given their beliefs[9], they have non-strategic reasons to use the rules they use. The context of interaction encountered by such a society does not amount to a coordination game as there is no alternative that they are inclined to follow, and hence Lewis (1969) must characterise their behaviour as non-conventional.

I take it we would still use the notion of 'convention' to describe the driving behaviour of the people in the first scenario, though not in the second scenario. This may be arguable in the case of those who made a promise to drive on one side of the road, but, in the case of those who follow linguistic rules in virtue of divine *fiat*, we would not hesitate to say that the community in question employs linguistic conventions. Matching our linguistic intuitions is not, as was explained before, an overriding goal of this inquiry. The deeper goal is that of trying to explain various forms of behaviour. Here the Coordinating rule view seems to carve up reality in a more useful way than Lewis' (1969) view does. It brings to the surface what we have in common with those who promise to drive on a specific side of the road, namely that this practice achieves a shared goal and does so to the degree that it is shared behaviour. A society may follow linguistic rules on religious grounds, yet their practice allows them to communicate for the same reason that we can communicate, namely that we all follow the same rules. Ultimately, it would be useful to be able to say, whether we follow linguistic rules on religious grounds or not, that some claim like '"Quine" refers to Quine' is true in virtue of a convention that is followed by users of 'Quine'. On the coordination view of conventions this remains true, even if some idiosyncratic society would rather stay quiet than use 'Krugman' to refer to Quine. In this way, the Coordinating rule view captures what such behaviour has in common, irrespective of the fact that their behaviour may be overdetermined[10].

4.4 Objection 4: Against characterising conventions in terms of exhibited regularities.

Lewis (1969) requires that every, or almost every member of a community must conform to the regularity involved. This is a strange thing to say, as presumably it is uncontroversial that a

---

[9] One could defend the Lewisian view by requiring that the beliefs in question must be true. In which case, stipulate that the society contains people who flog those who break linguistic rules.

[10] Lewis (1975) would object that, in effect, the coordination view allows games with one coordination equilibrium to count as generating a convention. His objection is that such games are trivial, as common knowledge of rationality is sufficient to a unique solution (Lewis 1975: 16–17). This is true, but I see no reason why we would demand that parties to a convention need to have a particularly challenging problem to solve. Especially if this leads to the view that punishing linguistic mistakes makes language less conventional, that laws punishing driving on the wrong side of the road makes driving on a specific side less conventional, and so on.

convention can exist in a community even if several members of the community violate it or are entirely unaware of its existence. In later work, Lewis (1976) clarifies the content of his theory by saying that a convention 'holds in a population' is ambiguous. It can mean that all (or almost all) of a group participate in a convention, or that some sub-population of the relevant population follow a convention (Lewis 1976: 116). The latter use, then, is the same as my use of the phrase 'sub-group of a population'. On the first usage, he would describe a situation in which there is non-universal conformity, and his other requirements are only partially met, by saying that in such a case a community has a convention "to a certain degree" (Lewis 1969: 78–80). The basic idea is that, if there is non-universal conformity, the community only has the convention to the degree that the relevant conformity obtains and the other criteria that he lists are met.

On both formulations though, the fact that conventions are defined in terms of exhibited regularities leads to the following oddity: imagine a community of a hundred people in which all are trying to drive on the same side of the road. Stipulate that the community lives on a large piece of land and that they do not drive very often at all. In fact, in this community one could drive on the wrong side of the road for quite some time without being alerted to one's mistake, both in virtue of not encountering other cars and, when encountering cars driving on the side different from one's own, assuming the mistake lies with them. Assume that a convention of driving on the left has been established, either in virtue of explicit agreement or spontaneously in virtue of the context of interaction being a coordination problem. Now imagine that, after some time, five people suffer a cognitive glitch and misremember the content of the convention. They now believe that the convention advises them to drive on the right, and they proceed to do so over an extended period of time. When they encounter someone driving on the left, they simply assume that the other person got it wrong. In such a case, it would be uncontroversial to say that there is one convention, namely driving on the left, and that the five deviants are party to the convention, even though they violate it. However, Lewis (1969) cannot portray the situation in this way. Given that Lewis (1969) defines conventions in terms of exhibited regularities, he would have to say, on the first usage, that the convention to drive on the left now exists to a lesser degree. Or, on the second usage, that the sub-population within which the convention exists has shrunk. Surely, portraying the situation in this way is perverse. We would never say that the convention now only holds to a degree, or that it now exists between fewer people. Lewis' (1969) view departs from common usage[11] and seems to miss something important, namely that the people involved are trying to coordinate, but some are simply failing to do so. The Coordinating rule view yields the intuitively compelling answer. All the relevant parties are disposed, absent defeaters and absent false beliefs, to drive on the left. Hence the Coordinating rule view yields the answer that there is a convention, namely to drive on the left, and that all one hundred people are party to it.

Note that the Lewisian analysis would also run into similar trouble where people's behaviour does not exhibit the required conformity due to the occasional presence of internal or external defeaters. On the Lewisian view this would undermine the existence of the convention to some degree. On the Coordinating rule view, and here it agrees with our common understanding of the situation, such cases do not affect the existence of the convention. The basic problem with

---

[11] The situation may be even worse. Lewis (1969: 64–68) states that the beliefs of the participants need not have general content, but need only be beliefs about the behaviour of specific drivers. If this is allowed, the five drivers can have the relevant attitudes, and so their behaviour and attitudes constitute a differing convention. Now the situation becomes one where there are *two* conventions, and these exist to radically different degrees.

characterising conventions in terms of exhibited regularities is that the required regularities may well not be exhibited, and yet the existence of the convention need not be affected in the least. If a German were to drive on the right in the UK due to a false belief about the dominant standard, or I swerve into an empty right-hand lane to avoid a pothole (external defeater) or to allow a police car to pass (internal defeater), this has zero consequence for whether a convention to drive on the left exists or whether I am party to the convention. This is so, even if the defeaters and false beliefs occur quite frequently. The Coordinating rule view captures this.

The above argument is consistent with the idea that, if a *sufficient* number of people were to start acting in accord with a deviant rule, the originally deviant rule may become the new convention. It follows from the nature of coordination games that a rational agent would be disposed to switch their allegiance if enough other people were to switch their allegiance, even if this originally occurred by mistake. The objection to Lewis's (1969) discussed above only pertains to cases where this has not yet happened, i.e. cases where the amount of people acting in accord with a deviant rule does not yet rationally motivate individual agents to switch their allegiance to the deviant rule. Lewis's contention that such cases are cases where the convention exists to a lesser degree, or that the sub-population in which the convention holds has shrunk, mischaracterises what is really going on.

5. Conclusion

In this paper, I have argued that Lewis' (1969) analysis of convention suffers from certain defects that can be resolved by adopting the coordination rule view. The problems are: (i) that it requires agents to understand their own situation too well; (ii) that it robs conventions of explanatory force; (iii) that it mischaracterises cases where someone has non-prudential reasons to follow a convention; and (iv) that it mischaracterises situations where the relevant behaviour is non-uniform. Characterising conventions as Coordinating rules, and the existence of conventions in terms of the conditions under which we can say such rules are followed, allows us to answer these objections and give a useful statement of the sense in which natural languages are conventional.

References

Blackburn, S. 1984. *Spreading the Word*. Oxford: Clarendon Press.

Burge, T. 1975. On Knowledge and Convention. *The Philosophical Review* 84: 249–255.

Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge: Harvard University Press.

Lewis, D. 1975. Languages and Language. In K. Gunderson (Ed). *Minnesota Studies in the Philosophy of Science*, Volume VII, Minneapolis: University of Minnesota Press. pp. 3–35.

Lewis, D. 1976. Convention: Reply to Jamieson. *Canadian Journal of Philosophy* 6: 113–120.

Miller, S. 1992. On Conventions. *Australasian Journal of Philosophy* 70: 435–444.

Schelling, T. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.