

A MORE RADICAL HUME

TAMÁS DEMETER

University of Miskolc, Department of Philosophy

In this paper I will discuss Hume's theory of convention that provides the background for his discussion of justice and the social contract. Conventions for Hume are socially held norms that govern our behaviour as members of societies. Hume's theory is reductive: the explanation of convention proceeds in naturalistic terms, in accordance with the requirements of methodological individualism. Being naturalistic in this sense means that it accounts for normative powers in terms that are not normative themselves, in terms describing properties that belong to us as normally functioning human beings and not as members of society. I will argue that his theory can and should be made more radical insofar as its naturalism is concerned.

One may reconstruct two accounts of convention in Hume, one may call them robust and minimalist accounts respectively (a similar distinction is drawn by Collin 1997: 199). The *robust* account appeals to the declaration of mutual interests that serves as a basis for regularities of conduct. These regularities are what he calls convention: „This convention is not of the nature of a promise; for even promises themselves, as we shall see afterwards, arise from human conventions. It is only a general sense of common interest; which sense all the members of the society express to one another, and which induces them to regulate their conduct by certain rules. I observe, that it will be for my interest to leave another in the possession of his goods, provided he will act in the same manner with regard to me. He is sensible of a like interest in the regulation of his conduct. When this common sense of interest is mutually expressed, and is known to both, it produces a suitable resolution and behaviour. And this may properly enough be called a convention or agreement betwixt us, though without the interposition of a promise; since the actions of each of us have a reference to

those of the other, and are performed upon the supposition that something is to be performed on the other part.” (Hume 1740: 490)

On the other hand, the *minimalist* Humean account does not require the declaration of interests. Convention can be formed in virtue of pure behavioural regularity among the parties to the convention. This lesson can be drawn from Hume’s famous example of rowing, where two people behave as if they agreed on a rate of strokes without actually making an agreement, or even expressing preferences to synchronise the number of strokes (Hume 1740: 490). On the weaker interpretation of the minimalist account, though mutual interests are not expressed explicitly, the parties to the convention nevertheless recognise them, and act in accordance with them. Thus conventions are built around the concept of interest. On a possible stronger interpretation, a sense of interest need not be present as constitutive to the convention, though they can be imposed on similar situations on a somewhat abstract, descriptive level.

Both Humean accounts represent an alternative to an explicit, agreement based, or contract-like account of convention. This is fairly obvious in the minimalist case, but seems to be true in the robust case, too. The declaration and/or recognition of mutual interests does not entail that the agreement itself needs to be of explicit nature. It is not presupposed in either picture that a social contract must be made in order for the convention to come into force. Thus Humeanism is in opposition to social contract theories like that of Hobbes (cf. Hume 1740: 493). These two traditions conceive differently the process of convention formation. Humeans are inclined to accept an evolutionary view where conventions appear from behavioural regularities, combined and built upon each other, and thus become even more complex as time goes. An outcome of this process is an institutional structure. This process is predominantly tacit, or implicit, even if it is based on mutual recognition of interests. For a Hobbesian the case is quite different. Institutions are formed through social contracts whose parties agree explicitly to create an institution for certain purposes.

Hobbesianism and robust Humeanism rely inherently on language. Social contracts cannot be agreed upon, and interests cannot be expressed without using language. Hobbesianism, of course, is stronger in this respect than robust Humeanism as the act of making contract presupposes certain verbal features. Though David Lewis (1969: 34) allows declaration of intention to be sufficient for reaching agreement without promises, it is not plausible that any agreement can be reached without some sort of explicit

commitment, i.e. without something that is promise-like. Giving promises is constitutive in making contracts, as the parties must commit themselves to obey contracts (Searle 1995: 35, adopts a similar position). One cannot make promises without using language, but for Humean promises themselves arise from conventions (Hume 1740: 490), thus they, and contracts along with them, are not eligible to provide the general basis of conventions. Robust Humeanism, as we have seen above, requires only the explicitness of interests, and does not rely on language in general, and promises in particular, in the process of convention formation. Minimalist Humeanism does not need any linguistic components in convention as it relies exclusively on behavioural regularities. Thus it becomes possible to treat language itself as an institution evolving gradually from convention (Hume 1740: 490). Lacking constitutive linguistic elements in a theory of conventions, this perspective can account for conventions that are not dependent upon contract or explicit mutual agreement.

Amending a Humean account of convention is best to be done along Humean lines by reinterpreting the requirements a Humean account poses for convention. As I mentioned above, these requirements are naturalism and methodological individualism. First I will take a closer look at naturalism, and then at methodological individualism. The function of Hume's theory of convention is to provide a naturalistic basis for moral sentiments, preferences and motives. Naturalistic in this sense means an account of normative power in terms that are not normative themselves. Humean conventions prescribe the way we should behave in certain circumstances, they are rules or sets of rules to which one can conform or refuse to conform. The naturalistic strand in Humean arguments tries to account for this normative power in a naturalistically acceptable, i.e. non-normative way.

David Lewis' account of convention meets the two criteria of Hume's model. First, a Lewis convention can serve as a reductive basis in the first step. Conventions may be a species of norms, namely regularities we think one is supposed to conform. If one fails to conform, then he goes against the expectations and preferences of others, and thus everyone else will infer that he knowingly acted this way. Therefore, he will face punishment. Second, Lewis' account is both naturalistic and individualistic

in the Humean sense. Though a Lewis convention can give rise to norms, it remains naturalistic because norms do not occur in the definition of convention.

For our present purposes we need a stronger sense of ‘naturalism’. Explaining the normative powers behind conventions may make use of a mentalistic vocabulary without violating Humean naturalism. But it has been questioned from two angles whether this allowance is legitimate. First, Paul Churchland and other eliminativists put forward an argument to the conclusion that mental states are nothing but posits of a deeply mistaken folk theory. Given that the theory is mistaken, its posits cannot be real, meaning among other things that they cannot be rightly invoked in the explanation of convention. Secondly, and more recently, Martin Kusch argued that mental states are social constructs in the sense that it is people as members of communities who create them. Mental states in this sense do not belong to us as normally functioning human beings, but as members of communities, i.e. they are not natural in Hume’s sense. Therefore we had better offer an analysis of convention in a more radical sense of ‘natural’, and this more radical sense should mean non-mental. In order to reach this result, I will combine two approaches to convention that are naturalistic in the sense required here. The first of these is evolutionary game theory that can provide a naturalistic translation of Lewis’ account. The second is Ruth Millikan’s proposal that emphasises reproduction and the weight of precedence in providing a naturalistic account of convention.

Evolutionary game theory is applied on the phenotypic level to decide the fitness of a particular phenotype if its fitness depends on others, and/or on the frequencies of particular phenotypes in a population. Standard assumptions of game theory are clearly out of place in this context, as they are useless in describing interactions among lesser animals: explaining the behaviour of reed warblers by invoking rationality may seem a bit strange, and the conceptual resources of a mentalistic idiom may be too generous in allowing the attribution of mental states to creatures that do not have the capacity to entertain them. Therefore, appropriate naturalistic concepts should replace the crucial mentalistic ones: rational conformity to a convention is to be replaced by evolutionary stability of a behavioural phenotype in a population, and the criterion of self-interest is to be substituted by fitness in these contexts. Pursuing self-interest in game theoretical contexts will be replaced by maximizing fitness in evolutionary game theory. This goes without saying that, fitness is maximized

consciously. Rather, it is maximized by way of selection: a behavioural phenotype with higher fitness results in more offsprings that may inherit and spread the same behavioural phenotype.

A behavioural phenotype can be understood as a strategy that specifies what an individual will do in certain circumstances. (Maynard-Smith 1982: 10) This strategy is evolutionarily stable if no mutant strategy can invade the population under the pressure of natural selection once all members of the population adopted it. Evolutionary stability can be shown in two ways: a) strategy₁ is evolutionarily stable if played against itself, its fitness is greater than the fitness of strategy₂ played against strategy₁; or b) strategy₁ and strategy₂ are equally fit against strategy₁ but strategy₁ is fitter against strategy₂. (Maynard-Smith and Parker 1976) Basic cases of convention can be treated as evolutionarily stable strategies. Consider a favourite example of Lewis, namely driving conventions. Let strategy₁ be driving on the left hand side of the road, and strategy₂ be driving on the right hand side. Suppose that strategy₁ is adopted by all members of the population. The fitness of strategy₁ in this case is greater if it is played against the conventional strategy, i.e. itself, than that of strategy₂ in the same case. Trying to invade a population by strategy₂ results in immediate collision or, if clusters of strategy₂ try to do so, in a series of collisions. This is a case of evolutionary stability in the first sense above. It is of course always possible to agree on replacing strategy₁ by strategy₂ but this would not cast doubt on the evolutionarily stable status of strategy₁, as this replacement would not be a result of the pressure of natural selection on strategy₁.

An evolutionary stable strategy produces an equilibrium state: no one is better off if any one follows a different strategy. And this is what we need for solving coordination problems by conventions. But solving coordination problems is impossible without cooperation: everyone must play one's own part in order to reach an equilibrium state. We have seen in Lewis' account how coordination equilibria can appear with the required presuppositions of rationality, mutual expectations, and so on. But how can equilibrium states appear without relying on preferences, expectations, rationality, etc., i.e. without using a mentalistic idiom? The answer is the Tit-for-Tat (TFT) strategy. In a population where iterated prisoner dilemma games are played TFT is the fittest of all strategies. Though TFT scores less against the Always Defect (AD) strategy, it compensates its losses against other TFT's. Though AD is an ESS against any strategy trying to invade the

population individually, still it can be invaded by a cluster of TFT's that may appear in the population easily due to kinship relations. Playing TFT does not require mind or consciousness. It requires only an ability to recognise different partners, or the special case of having one partner for a life. TFT can be a behavioural phenotype inherited via imitation of successful predecessors, or by trial: if everyone plays TFT, it does not pay playing something else. (Maynard-Smith 1982: 171f.) Similarly, successful behavioural phenotypes can be transmitted this way.

This explains how a strategy can be evolutionarily stable, but we still do not know which strategies can count as conventions. For Millikan conventions are composed of reproduced patterns and the proliferation of patterns due to the weight of precedence. A behaviour on this account is conventional not because it fits a conventional pattern, but rather because it was reproduced. This means that the conventionality of a behavioural phenotype does not require conformity, but merely that a given phenotype be derived from a previous phenotype by the reproduction of the relevant respects. In order for a behavioural phenotype to count as a convention it must be reproduced due to the weight of precedent, as opposed to some intrinsic capacity of the pattern to serve certain functions. Thus the conventionality of a pattern entails that it would not appear in the absence of precedent, and therefore a conventional pattern is arbitrary in the sense that it has alternatives that can substitute it equally well.

Consider then driving conventions again. How can we explain the emergence and maintenance of the convention of driving on the left hand side without relying on a mentalistic vocabulary? First, we need a precedent, or more precisely, several successful precedents in solving the coordination problem. Second, this conventional pattern needs to be reproduced due to the weight of precedent: by imitating successful predecessors, or learning from trial. This solution produces an equilibrium where everyone's fitness is higher if everyone behaves the reproduced way. Thus appears an ESS which correlates the behaviour of the parties to the convention, and which cannot be invaded by way of natural selection by another behavioural phenotype.

The basic notion around which the Humean account is built is rationality. Conventions are based on rational expectations about others' behaviour, and on rational preferences, based on the knowledge of the pay-off matrix, about

others' behaviour. These are also accompanied by rational behaviour on the agents' part guided by beliefs and desires that are formed according to these expectations and preferences. But what is the basis of the expectations and preferences? It is salience based on precedents. Can we say then that they are rational? We cannot say this because they appear due to a process of induction: on the basis of previously successful behaviour the agents infer that other agents will follow the previously successful procedure. However, as it should be clear to a Humean, the process of induction can't give rise to rationally justified expectations. Past success does not guarantee anything about others' behaviour and therefore it is not rational to form expectations along these lines. Preference as a basis of convention cannot cope with rationality as a pillar of a mentalistic account in folk-psychological terms.

Without making use of rationality it is impossible to give a mentalistic account of convention. If the parties to the convention are not rational then whatever expectations, preferences, and knowledge they may have, they will not conform to the convention, their behaviour will be random. But given that conventional conformity is based on precedent it cannot be based on rational expectations, therefore the folk-psychological conceptual framework must be abandoned.

The place of rationality in an account of convention is precarious from the perspective of cooperation, too. Accepting a Humean methodological individualistic notion of rationality that appears in Lewis' game-theoretical account makes it dubious that coordination can appear in prisoner dilemma games. For example, the highly successful TFT strategy could not appear, or become evolutionarily stable in a rational community. If agents are rational, then none of them will cooperate in the first step because in the case of defection from the other's part she could never regain the loss she suffered in the first step. Considering iterated games, rationality would fail to explain the maintenance of a cooperative strategy, even if successful before, due to induction on the one hand, and to the promise of an immediate better pay-off in the case of defection on the other. Thus rationality falls short of explaining the relevance of precedence for convention, and the possibility of cooperative conventions in prisoner dilemmas.

The study can be summarized as follows: Rationality must be the unifying concept behind any explanation of convention conforming to the requirement of Humean methodological individualism. Without rationality the conceptual resources of a mentalistic idiom are not accessible for an

account of convention. However, rationality cannot explain the role of precedence and the maintenance of cooperation, therefore it cannot explain convention. If rationality cannot explain convention, then the mental cannot play a role in explaining convention in the first place. Therefore a holistic or a methodological eliminativist account is in order that avoids using mental predicates. Due to Occam's razor, however, the latter seems more favourable.

References

- Churchland, Paul 1981: „Eliminative materialism and the propositional attitudes”, *Journal of Philosophy* 78.
- Collin, Finn 1997: *Social Reality*, London: Routledge
- Hume, David 1740: *A Treatise on Human Nature*
- Kusch, Martin 1999: *Psychological Knowledge*, London: Routledge
- Lewis, David 1969: *Convention*, Cambridge/Mass.: Harvard UP
- Maynard-Smith John 1982: *Evolution and the Theory of Games*, Cambridge: UP
- Maynard-Smith, John – Parker, G.A. 1976: „The logic of asymmetric contests”, *Animal Behaviour* 24, 159-175.
- Millikan, Ruth 1998: „Language conventions made simple”, *Journal of Philosophy* 95, 161-180.
- Searle, John 1995: *The Construction of Social Reality*, New York: The Free Press