PLEASE SCROLL DOWN FOR ARTICLE

The Enigma of Rules[*]

Jaroslav Peregrin

Abstract

In a remarkable early paper, Wilfrid Sellars warned us that if we cease to recognize rules, we may well find ourselves walking on four feet; and it is obvious that within human communities, the phenomenon of rules is ubiquitous. Yet from the viewpoint of the sciences, rules cannot be easily accounted for. Sellars himself, during his later years, managed to put a lot of flesh on the normative bones from which he assembled the remarkable skeleton of the early paper; and his followers too. However, what they say is somewhat divergent; and therefore my aim in this paper is to concentrate on the very concept of rule and analyse it in the context of the question what it is about us humans that makes us special.

Keywords: rules; normativity; evolution; language

In his remarkable early paper 'Language, Rules and Behavior' (1949), Sellars warns us that if we cease to recognize rules, we may well find ourselves walking on four feet. This sounds sufficiently menacing to spur us on to try to understand what is so special about rules and why they are so essential to us humans. Sellars himself, during his later years, managed to put a lot of flesh on the normative bones from which he had assembled the remarkable skeleton of the early paper; and his followers too (not only the 'left-wing Sellarsians' like Brandom or McDowell, but also some of the 'right-wing' ones, like Millikan) have much to say about norms and rules. However, what they say varies somewhat; and therefore my aim in this paper is to concentrate on the very concept of a rule and analyse it in the context of the question of what it is about us humans that makes us special.

It is obvious that within human communities, the phenomenon of rules is ubiquitous. We have the all-important rules that are codified by our law, we have rules that are not officially written down but are usually followed (like the rule that if somebody helps me, I should be prepared to help him in turn), we have rules about traffic and the rules of various games and sports. Yet from the viewpoint of the sciences, rules cannot be easily accounted for. How are we to explain their emergence (compatibly with evolution), and how are we to account for their mode of existence, especially where they are

unwritten? Are we to identify a rule always with some kind of *linguistic* object; or are we to reduce it to a regularity of behaviour?

Neither of these two options seems promising. There are certainly rules that exist without being recorded and hence without there being any linguistic object with which they could be identified. (After all, we talk about the *encoding* of the law, which seems to suggest that the law articulates something that is here independently of the code.) And reducing the existence of a rule to a certain kind of regularity of behaviour would seem to wipe out any distinction between billiard balls 'following the rules' of mechanics and human subjects following the rules of their society. Hence it would seem that though there must be more to rules than regularities, at least some rules must be capable of existing exclusively 'within' human conduct – being, as Sellars (1949: p. 299) put it, written 'in flesh and blood, or nerve and sinew, rather than in pen and ink'.

Rules and Evolution

Currently it seems that what can help us get a grip on the concept is an inquiry into the ways rules manage to come into being. And indeed, evolutionary biologists are nowadays preoccupied with phenomena that seem intimately connected with rules, namely with the phenomena of cooperation and altruism. How is it possible that people do things that seem to be beneficial for their peers rather than for themselves? How do they come to bind themselves with rules that may sometimes make them diverge from the trajectory dictated to them by their apparent needs, the following of which natural selection hammers into their genes?

Several answers to these questions have already been proposed. Once Dawkins (1989) had convinced his colleagues to replace *individual* in the centre of the evolutionary picture by *gene*, the explication of altruism with respect to ones kin was forthcoming. Meanwhile, several biologists have tried to explain other versions of altruism as a matter of *tit-for-tat* (Trivers, 1971; Axelrod, 1984, 1986). We do good to our peers, they propose, because we have reasons to believe that they will do good to us later, and that the final account will be profitable for us. Hence the idea is that altruism is a profit-making investment.

Now imagine two creatures, *A* and *B* (call them *hunters*) confronting each other over a killed animal, whose meat amounts to, say, six energy units. Assume that each of the hunters may be disposed in either of two ways: to fight for the whole supply of meat or to resign the fight (see Figure 1). Altogether, then, there are four possible cases: if both go for a fight, each of them will, in the end, get his three units (assuming their physical dispositions are comparable and average out over multiple cases), but both will lose some energy through the fight, say two units. If only one of them is ready for a fight, whereas the other withdraws, the first will get the whole six units, but

A \ *B*	fight	resign (cooperate)
fight	*A*: 1, *B*: 1	*A*: 5, *B*: 0
resign (cooperate)	*A*: 0, *B*: 5	*A*: 3, *B*: 3

Figure 1.

being unable to consume them all at once, he will have to save part of the meat for the future, making his final energy gain less than six – hence, say, five – units (storing will cost some energy, and the storage itself may reduce the energy value of the meat). The withdrawing hunter, of course, will get nothing. If neither wants to fight, they may share the meat and each of them will get three units.

This suggests that from the global viewpoint, sharing would be the most profitable strategy, for it maximizes the gross number of energy units distributed among the members of the hunter community. The trouble is that from the viewpoint of an individual hunter, the situation looks different; indeed, from his viewpoint the unambiguously most profitable strategy is fighting. The reason is that if his peer wants to fight, fighting will secure him at least one unit (whereas withdrawal would secure him none); and if his peer does not, then fighting will secure him five units (whereas not fighting would secure him only three). In game theory (Maynard Smith, 1982), which models the situation just envisaged in terms of the so-called *Prisoner's Dilemma* (Poundstone, 1992), not fighting is what is called a *strongly dominated strategy* – whatever the opponent does, fighting turns out to be more profitable than not fighting.

Hence we may see the problem as consisting in the fact that a rule cannot be operative unless it is endorsed by many people. ('I would happily give a share of my meat to my comrade, if I knew that he would give me a share of his some time in the future, but how can I be sure?') From this point of view, the ideas of *reciprocal altruism* and *tit-for-tat* can be accommodated only if we change the settings, in particular if we assume that the dispositions of the hunters do not concern strategies with respect to individual encounters, but

rather to series of such encounters. (It seems that there is no reason to suppose that it could not be the entire series together that wields evolutionary pressure.) Also, of course, fighting and resigning are not the only available strategies – we could adopt *mixed* strategies such as 'start to cooperate, but go on cooperating only with those who reciprocate'. Some such strategies may be viable (see Lehmann and Keller, 2006 for an overview).

Besides these, there are other dispositions that may foster cooperation, i.e. make the member of the community stick to cooperating rather than fighting. One of them is the disposition towards so-called (*altruistic*) *punishment* ('chastise those who are not willing to cooperate' – Fehr and Gächter, 2002). Many theoreticians argue that starting on the journey to a stable social order as we know it from our communities requires more than becoming cooperative or altruistic (a community of cooperators is vulnerable to an invasion of 'parasites' who want to profit from cooperation without contributing anything themselves, as individuals with such devious, parasitic dispositions are always bound to appear, as a result of mutations). What is needed, in addition to cooperation, is punishment for those who are not willing to cooperate. Moreover, it seems that there might be a need for a third level of behaviour: not only to be altruistic oneself, and to make others altruistic too, but also to make others make others altruistic ('chastise those who are not willing to chastise those who are not willing to cooperate' – Heckathorn, 1989). Besides punishment, another significant factor may be selectiveness with respect to cooperative partners ('not only do not cooperate with those who do not reciprocate, but try to completely avoid them'). This creates special 'social networks' where cooperation may flourish (Woodcock and Heath, 2002).

These evolutionary stories are instructive and important, and we will return to them later. But at this point I want to suggest a change in visual angle. My conviction is that connecting the general idea of a norm or a rule too closely with the ideas of cooperation and altruism may be misleading – it may obscure another important role of rules. If we take a look at these matters from a less usual viewpoint, we may see an important aspect of the phenomenon of rules which is currently eluding us.

Let us notice that what the evolutionary stories explain is especially 'heavy-weight' rules, rules that have to do *directly* with our survival and the violation of which may cost us, if not directly our life, then at least something else that truly matters to us (these are the rules of the *moral* kind in the narrow sense – from 'You shall share your meat' to the legendary 'Thou shalt not kill!'). But what about the rules of, say, football? This question may seem preposterous. Are not the rules of football something utterly different from moral norms? It does not seem to be difficult to explain the existence of games and sports from the evolutionary perspective (a training for the struggle for survival …), but the emergence of games seems minor compared to the problem of the emergence of altruism!

However, the question is not why we have games, but why we have games *governed by rules*. (After all, children are happy playing without using any true rules, or at most rudimentary ones.) And what I want to suggest is that the difference between the rules of football and the rules of morals is not so serious that we could not try to see all these varieties of rules as species of a single kind. I think that asking the general question about why we have this very kind of institution might bring about the desirable stimulating change of visual angle. This suggestion is backed by the conviction that, though clearly there are many fundamental differences between morals and football (between, to put it in the form of an aphorism, 'You shall not kill!' and 'You shall not touch the ball with your hands!'[1]), they also have many important features in common.

So what essentially differentiates the rules of morals from those of football? There seem to be at least two fundamental differences: firstly, moral rules are incomparably *more important*, and secondly, whereas moral rules seem to be *categorical* (applicable unconditionally, i.e. in force always and for everybody), the rules of football are *hypothetical* (applicable only conditionally, i.e. in force only, for example, for those who choose to pursue some goal). As for the first difference, it seems indisputable that whereas the rules of morality lie in the very foundations of human sociality, the rules of games or sports concern something more parochial and dispensable. But though the difference is obvious, it is far less obvious that it cannot be construed as one of *degree* rather than of *kind*. (Some of the rules that we might classify as moral surely hold less importance for our present society than those of football.)

As for the second difference, again it seems clear that whereas moral rules are binding for everybody, the rules of football apply only to those who elect to be part of the game. But this difference is perhaps even less firm than the previous one. In fact, just as the rules of football delimit what it is to be a football player, the rules of morals delimit what it is to be a human being. We do not apply them to individuals of other species: a tiger killing an antelope is not considered as violating any moral rule. Moreover, and this is important, they need not be applicable even to all humans – if a group of biological humans were to live totally amorally (without in any way interfering with us), our decision might be simply not to consider them true members of the 'human race' and leave them alone. This indicates that the term *human* related to the principles of morals may not be a biological one, but one *constituted* by the principles of morals. Hence it would seem that it is not too far-fetched to say that just as the rules of football delimit the arena of football, so the rules of morals delimit the arena of humanity.

Once we see the differences between the various kinds of rules as not totally alien to one another, we can see the common core. Rules regulate human conduct – they are applicable only to creatures that we consider to have a free will. Something is a rule only in so far as those governed by it are

capable of doing otherwise than prescribed by it. Rules make people behave in certain ways – enforce behavioural patterns. How do they do this?

A behavioural pattern can be wired into a human brain (or, for that matter, into a brain of another animal) by natural selection. But this, of course, is not the only way that such a pattern may come into being. A person may become conditioned if she is rewarded when behaving in accordance with the pattern and penalized when not. Why would her peers do this rewarding and penalizing? Perhaps *they* have this 'normative' behaviour wired into their brains by natural selection. (Remember the concept of altruistic punishment.)

But this seems strange. Why would evolution enforce the pattern in such a roundabout manner, producing its 'enforcers' who force it upon 'enforcees' instead of making the enforcees display it right away? And would this kind of enforcement not lead to a selective advantage for those with an inborn adherence to the pattern, thus wiping out the others and soon enforcing the pattern directly after all?

Well, imagine that what the enforcers of the pattern would be capable of doing would be not only to make the enforcees display such a pattern, but also to make them make others display it – hence not only to become adherents of the pattern, but also its enforcers. If this were possible, the pattern would become capable of a purely 'cultural' promulgation, and would need no wired-in support. In this way, the promulgation of behavioural patterns standardly effected by evolution would bear another level of such promulgation, piggybacking on it but going its own way.

Culture

The idea that at some stage the standard genetic replication bears a higher-level, 'cultural' descendant (which, though piggybacking on it, may assume a pace and a trajectory largely independent of those of its carrier) is surely not a new one. In Dawkins' path-breaking book about evolution (1989), it received a suggestive shape centred around the concept of *meme*; and it gave rise to the proposal that memes, the cultural analogues of genes, are replicated by imitation, fighting for survival in their abstract milieu just as genes fight for their survival in their concrete one.

The basic picture, then, is that the working of the fundamental replicators, genes, gives rise to a different, 'higher-order' form of replicators, *viz.* the memes. Why does this happen? Well, Dawkins' answer seems to be that though memes are not much more than a by-product, they are capable of providing for a certain surplus in survival value: 'Once the genes have provided their survival machines with brains that are capable of rapid imitation, the memes will automatically take over.' Hence memes, though originating as mere 'spin-offs', are able – at least to a certain extent – to steal the show. What seemed to be the crux of evolution, genes

devising more and more complicated vehicles of individual bodies with their brains, etc., for the purposes of their self-promulgation, now seem relegated to the maintenance of a medium in which another kind of evolution sets off.

Dawkins' basic principle, which makes the replication of memes possible, is *imitation*. According to him, it is because one individual is capable of imitating another one that the memes can start spreading across human societies, thus being able to interact with each other, compete and fight for survival, just like genes. But this is where, I think, Dawkins' story loses some plausibility. Should we really believe that the path-breaking change of the course of evolution, the creation of, as it were, its brand new layer, is a matter of *imitation*?

It would seem that what makes us humans unique, what makes our antics, in contrast to those of other species, deserve the specific name of *culture*, is precisely that we are able to go *beyond* imitation – we do not *copy* ideas (memes) of our peers, we engage in very complicated interactions in the course of which the 'memes' get upgraded. Dawkins tries to account for this in terms of *imperfections* in the way we copy memes – people, according to him, often do not quite imitate one another, but do it only imperfectly. (Thus, Dawkins, for example, claims to replicate, in his book, some memes of other authors, but to replicate them imperfectly, by which he means that he does not merely repeat them, but elaborates on them and advances them.) But this sounds rather odd: at the very least it seems that *imperfection* is a very inadequate word to characterize the difference between mere imitation and the way of upgrading which is really going on.

Moreover, it seems that it is not adequate to see the upgrading as a matter for an individual. Upgrading ideas is usually teamwork, and ever more so. This is not to say that to get upgraded an idea must change hands more than once; it is to say that memes are essentially distributed. They do not exist via individual humans, but via networks of human interaction within human societies.

But here the modifications of the theory of memes that would be necessary to make it realistic start to make it too complicated to be workable. Therefore, I propose approaching the matter from a wholly different angle, leaving the problematic concept of a meme behind. Instead, I propose to concentrate on the general concept of a rule. What I suggest is that we look at the surplus, 'cultural' level of evolution that Dawkins rightly pointed out not in terms of the concepts of *imitation* and *meme*, but in terms of the concept of *rule*. (However, I think that the new level should not be seen as providing directly for evolution in the Darwinian sense, but rather merely for a more effective accomplishing of a task that was being accomplished by evolution before – namely the spreading of behavioural patterns.)

383

Sellars on Rules and Pattern-Governed Behaviour

Many philosophers during the last half-century have addressed the concept of a rule. A meticulous philosophico-logical analysis of the concepts of *norm* and *rule* was given, for example, by von Wright (1963). But here I want to concentrate on the analysis given by Sellars, which I will try to show may be seen as surprisingly relevant for the evolutionary formulation of the problem.

Sellars (1954) realized that our language games provide for an example of an activity that is neither *merely conforming to rules* ('doing *A* in *C*, *A′* in *C′* etc., where these doings "just happen" to contribute to the realization of a complex pattern') nor fully fledged *obeying of rules* ('doing *A* in *C*, *A′* in *C′*, etc. with the intention of fulfilling the demands of an envisaged system of rules'). On the one hand, language games would fall into the same category as any regular happenings, such as things falling in conformity with the law of gravity or planets circling the sun in their normal manner; which seems to be simply unacceptable. However, on the other hand, assuming that any linguistic action presupposes a comprehension of some rules would lead us to a vicious circle, for we would have to comprehend the corresponding rules *correctly*, and hence would need to follow rules of interpretation.

This led Sellars to stipulate a middle way between the two extremes; he urges that in between the rule-conforming and the rule-obeying behaviour there is another important kind of behaviour, which he calls 'pattern-governed'. This kind of behaviour is not like the merely rule-conforming one, for there is a sense in which we can say that it is done 'because of the system', but on the other hand it is not like the rule-obeying behaviour, for it does not involve an explicit comprehension of the system. Sellars (1954: pp. 207–8) gives two examples of such behaviour, i.e. of behaviour that is done because of a system, but not because of the comprehension of the system; and both concern evolution.

The first example runs as follows:

> In interpreting the phenomena of evolution, it is quite proper to say that the sequence of species living in the various environments on the earth's surface took the form it did because this sequence maintained and improved a biological rapport between species and environment. It is quite clear, however, that saying this does not commit us to the idea that some mind or other envisaged this biological rapport and intended its realization. It is equally clear that to deny that the steps in the process were interrelated to maintain and improve a biological rapport, is not to commit oneself to the rejection of the idea that these steps occurred because of the system of biological relations which they made possible. It would be improper to say that the steps 'just happened' to fit into a broad scheme of continuous adaptation to the

environment. Given the occurrence of mutations and the facts of heredity, we can translate the statement that evolutionary phenomena occur because of the biological rapport they make possible – a statement which appears to attribute a causal force to an abstraction, and consequently tempts us to introduce a mind or minds to envisage the abstraction and be the vehicle of its causality – into a statement concerning the consequences to particular organisms and hence to their hereditary lines, of standing or not standing in relations of these kinds to their environments.

(Sellars, 1954; pp. 207–8)

The second example follows:

What would it mean to say of a bee returning from a clover field that its turnings and wigglings occur *because* they are part of a complex dance? Would this commit us to the idea that the bee *envisages* the dance and acts as it does by virtue of intending to realize the dance? If we reject this idea, must we refuse to say that the dance pattern as a whole is involved in the occurrence of each wiggle and turn? Clearly not. It is open to us to give an evolutionary account of the phenomena of the dance, and hence to interpret the statement that this wiggle occurred because of the complex dance to which it belongs – which appears, as before, to attribute causal force to an abstraction, and hence tempts us to draw upon the mentalistic language of intention and purpose – in terms of the survival value to groups of bees of these forms of behavior. In this interpretation, the dance pattern comes in not as an abstraction, but as exemplified by the behavior of particular bees.

(Sellars, 1954; p. 208)

Finally, Sellars gives us direct instructions as to how to apply these evolutionary examples to 'the phenomena of learning':

Indeed, it might be interesting to use evolutionary theory as a model, by regarding a single organism as a series of organisms of shorter temporal span, each inheriting disposition to behave from its predecessor, with new behavioral tendencies playing the role of mutations, and the 'law of effect' the role of natural selection.

(Sellars, 1954; p. 208)

This instruction, as it stands, may be puzzling, for what is essential for selection is competition among an abundance of alternatives, whereas Sellars

speaks merely about a succession of organism-stages; but I think that it is not difficult to see what Sellars has in mind. Obviously, what he means by 'regarding a single organism as a series of organisms' is seeing an organism as a trajectory over an often branching tree of possibilities concerning behavioural patterns. At each point, only one kind of pattern, the one most appropriate to the pressures of the environment, survives and then gets us to the further branching point with further possibilities of its further development.

What is going on, then, is the selection of certain behavioural patterns from an offer of many possible alternatives; a selection which, in the end, allows us to say that the organism makes or does something because of the pattern, but not because of its comprehension of the pattern. How does this selection proceed? Of course, by the coercion of the teachers; but this coercion is the result of a specific dialectic, the dialectic of what Sellars (1969) calls 'ought-to-do's' and 'ought-to-be's'. *Ought-to-do*'s are simply commands, prescriptions that an agent is to do so and so. To comprehend them, the agent has to possess relevant concepts, concepts which make up the *ought-to-do*'s. They may be thought of as imperatives. *Ought-to-be*'s, in contrast, are not construable as commands, as they are not explicitly directed at an agent. Rather, they mark a state as desirable. They *may* lead to actions, because they bear *ought-to-do*'s, via a specific kind of generic 'practical syllogism': *If something ought to be, and doing A is likely to bring it into being, then do A*. Again, one must comprehend the relevant concepts to use the *ought-to-be* to follow this syllogism.

But in addition to being an *agents* following *ought-to-do*'s and endorsing *ough-to-be*'s, a person may also be a *subject* of an *ought-to-be*. And, according to Sellars, there is a serious difference between *X should do A*, which requires *X* to be an agent comprehending *A*, and *X should be in state φ*, which does not involve any such requirement. The latter is rather a 'free-floating' norm, which is up for grabs for *any* agent and comprehender (including, possibly, *X* herself). And Sellars' claim is that language learning is moving from the position of a subject of certain *ought-to-be*'s to the position of their endorser:

> [T]he members of a linguistic community are *first* language *learners* and only potentially 'people,' but *subsequently* language *teachers*, possessed of the rich conceptual framework this implies. They start out by being the *subject-matter* subjects of the ought-to-be's and graduate to the status of agent subjects of the ought-to-do's. Linguistic ought-to-be's are translated into *uniformities* by training.

> (1969: p. 512)

This indicates that somehow bringing into being an *ought-to-be* of the form *One should be in a state φ*, by teachers of somebody *X*, forces, in the long

run, not only that *X* be *in state φ* (via the commands of the form *Y, act so as to put X in state φ*!, which the teachers derive from *X should be in state φ*, which is in turn derived from the original *One should be in state φ*), but also *X*'s comprehension of the ought-to-be (and consequently deriving commands of the form *X, act so as to act Y in state φ*). In short, when educating humans (or candidates of humanity), enforcing behavioural patterns results not only in the patterns' coming into being, but also in the patterns, being endorsed as an ought-to-be.

How can this happen? Well, we may conjecture that a human agent being forced into a preconceived pattern inevitably comes to reflect and represent the pattern; and comes to represent it as something that is desirable. Perhaps this can be seen as the biological correlate of us humans being 'normative beings' – we tend to understand a certain kind of coercion as a manifestation of an *ought-to-be*. This is what brings into being the evolutionary mechanism envisaged above – the enforcement that makes the enforcees become not only adherents of the pattern enforced, but also its enforcers.

This appears to be precisely what makes up, from the viewpoint of the behavioural patterns, a *rule*: a general and generic desideratum concerning members of a community, the implementation of which brings about the implementation of its desirability. In other words: certain ways of forcing you to do *A* rather than *B*, in certain circumstances, make you not only do *A* in the circumstances, but also construe *A*, and not *B*, as being *proper* in those circumstances, the consequence of which is that you will force others to do *A* rather than *B* in the circumstances. It is in this way that the rule comes to perpetuate.

If we now return to the game-theoretical models of cooperation, we can see that it is precisely this aspect of rules that makes for the factors diagnosed as crucial for the stabilization of cooperation, such as altruistic punishment. Once I take a state as desirable, I not only behave so as to bring about and sustain the state, but I also try to make others bring it about and sustain it.

Hence rules institute the very kind of circle that, as we indicated above, is *reproductive* in the sense that it provides for a kind of 'evolution in evolution' – for the 'cultural' spreading of 'software' behavioural patterns piggybacking on the 'natural' spreading of the 'hardware' ones. The relevant patterns are forced upon us not (directly) by natural selection, but by the continuing demands of our peers. A *rule* is a lever necessary for putting to work the exclusively human kind of forming and maintaining of patterns – it is 'an embodied generalization which to speak loosely but suggestively, tends to make itself true' (Sellars, 1949: p. 299).

Standalone versus Integrative Rules

Wittgenstein (1969: pp. 184–5) pointed out the distinction between two kinds of rules:

387

Why don't I call cookery rules arbitrary, and why am I tempted to call the rules of grammar arbitrary? Because I think of the concept 'cookery' as defined by the end of cookery, and I don't think of the concept 'language' as defined by the end of language. You cook badly if you are guided in your cooking by rules other than the right ones; but if you follow other rules than those of chess you are playing another game; and if you follow grammatical rules other than such and such ones, that does not mean you say something wrong, no, you are speaking of something else.

Rules of cooking – as well as many other rules of the same kind[2] – are determined by the end of cooking: to cook correctly simply means to prepare various kinds of edible and tasty meals. On the other hand, the rules of chess are not determined by the end of chess. In comparison to the previous ones they give us a dimension of freedom – there is nothing which would force us to accept a rule that bishops move diagonally analogously to how we are forced to accept the rule that meals should not contain too much salt!

Does this mean that in the case of rules of Wittgenstein's latter kind human freedom and human spontaneity come into the open? As a matter of fact I think that it does, but we should be careful not to misconstrue the situation. Does the arbitrariness of the rules of chess or of language mean that chess or language have no purpose? Does it mean, for example, that they have no evolutionary explanation?

I do not think that this is the case. I think that what is the case is that if there is an evolutionary explanation for either chess or language, then it is the explanation of the whole enterprise, not of the individual rules. Though any individual rule is arbitrary, what they make up together is no longer so. The arbitrariness derives from the fact that there may be many ways to do justice to the purpose of the whole thing – as the plethora of natural languages testifies, there are many equally good ways to accomplish what English or German or Turkish accomplish in their ways.

This institutes a crucial holism characteristic of this kind of rules. We have already encountered what could be called an 'interpersonal holism': 'a rule cannot be operative unless it is endorsed by many people'. (This kind of holism was responsible for the clash between the collective perspective, from which the rules of cooperation were unambiguously profitable, and the individual one, from which one always depends on the goodwill of others.) Here there is an additional dimension of holism, a kind of 'internormative holism': 'a rule cannot be operative unless it is endorsed together with many other rules'. Let us call the rules displaying this additional holistic dimension *integrative*.

This perspective, I believe, may throw some new light on the distinction between Brandom's theory of normativity and those theories that try to account explicitly for normativity in terms of evolution, such as Millikan's

(2004) *teleosemantics*. Millikan insists that any norm worth its name is a matter of 'natural purpose', of 'what a biological or psychological or social form has been selected for doing, through natural selection' (Millikan, 2005: p. 65). Dennett (2008), who appears basically to share this attitude of Millikan, duly points out that Brandom, in contrast to this, sees error not as a case of 'faulty design', but rather of 'social transgression' (Dennett adds: 'Roughly, it is the difference between being stupid and being naughty.') Does it mean that Brandom would want to see the norms as coming from something other than a natural development?

I do not think so (though I can understand Dennett's frustration at Brandom's totally ignoring questions concerning the source of the norms). I think that what should reconcile the views of Brandom and Dennett might be the admission, on Brandom's part, that language, as well as other integrated systems of norms, do have an evolutionary purpose, and the recognition, on Dennett's part, that such systems have a purpose as *wholes*, so that there is a sense in which individual norms are arbitrary – the holistic nature of the whole system enables it to be constituted in different ways. (This would converge to the thesis that some errors do amount to 'being naughty' rather than to 'being stupid', but we can always say why it would be stupid not to chastise people 'being naughty' in this way.)

Rules as Opening Virtual Spaces

We have seen that from the viewpoint of evolution, it is the 'heavy-weight' rules, especially rules of morals, that are crucial. Other rules, like the rules of football, can then perhaps be seen as their 'parochial simulacra' (football as 'morals of the playground') – we simply remove some weight of the moral rules and gain 'light-weight' rules that do not trouble us unless we are bored enough to want to play. And rules of language, though surely not so easily avoidable as those of football (we cannot help playing our language games), belong with the light-weight ones – it would be hard to lose one's head or one's property for not respecting the rules of English.

In other words, the usual way of thinking about rules and evolution is that at some point of evolution, 'altruism', 'cooperation' or 'collective action' became profitable and the emergence of rules is due to the fact that rules are somehow able to implement just this. But we have already suggested that what distinguishes the rules of morals from those of football may be less important than what these two kinds of rules share. Perhaps rules and altruism are not as intimately connected as we tend to think; perhaps what is crucial is not that rules allow us to cooperate and make reciprocal altruistic investments; perhaps the truly crucial thing rules bring us is something else.

Hence my suggestion, in the form of an aphorism, is that in the sense under discussion, football is no less basic than morals. Perhaps, that is to say,

light-weight rules are not secondary to the heavy-weight, moral ones. And as among the things that are driven by the light-weight rules we find language, the emergence of such rules would mean not only the possibility of playing prehistoric football, but also the possibility of talking. And this is not something that is in itself light-weight, even from the viewpoint of evolution.

But if cooperation is not the most basic achievement rules are responsible for, what is it? We have already given part of the answer: *rule* is a complex 'meta-pattern' that underlies the cultural spreading of behavioural patterns. It provides for patterns that can be passed down not only as such, but including the comprehension of their desirability, which causes them to be perpetuated. Let us now complete the answer.

Success in evolution is a matter of fitness *with respect to an environment*. (It is trivial that being fit with respect to one kind of environment may well be being unfit with respect to a different one.) Now, once our predecessors started to form communities, part of the relevant environment came to be constituted by their peers. (This led to the result that fitness may be a matter of certain equilibria rather than simply of an optimization of features.) Moreover, when communities started to function as what can be called societies (i.e. when rules started to play a crucial role), the tangible barriers of nature that channel evolution became increasingly replaced by artificial ones. We twenty-first-century Westerners evolve because of pressures that have little to do with the availability of natural resources or with fighting for survival with our own hands; the pressures that shape us now have to do with social standards and our abilities to live up to the needs of our society.

And what I want to stress is that it is rules which have led us to the establishment of 'virtual worlds' – virtual not in the sense of being unreal, but in the sense of owing their existence to the attitudes of us people, namely to our *normative* attitudes that sustain the integrative rules necessary to underpin such virtual edifices. In this way, rules provide for a basic alteration of the human niche and consequently of its evolution-fuelling features. And it is in this way, too, that rules provide for an acceleration of evolution, for they rob genetic replication of its exclusive right to promulgate patterns. Now we see the mechanism behind it in full plasticity: rules provide for evolution's self-adjusting of the barriers against which the selection that fuel it takes place.

Consider the development of computers. At first, the development ('evolution') was a matter of the improvements of hardware. But once the idea of a multi-purpose hardware appeared, a hardware that is not devoted to one pre-conceived task, but is rather versatile and can be adapted, via software, to cope with various kinds of tasks, the situation changed radically. It is not that the evolution of hardware has stopped, but that it is no longer guided directly by the tasks the computers are to cope with (the 'environment') – rather it is guided by the tasks to support, as well as

possible, the kind of software that is able to cope with the more basic tasks. And the 'front-end' layer of evolution is that of software – it is software that, though not able to exist without the hardware, faces the environment directly.

The metaphor of hardware and software is well known from the philosophy of mind – there it is usually the brain that is compared to the hardware and mind is thought of as the software (see, e.g., Block, 1995). But here I am employing it in a different way (of course, I am not claiming originality even for this metaphor): cultural evolution as software running on the hardware of the natural one. I think that this metaphor is much more realistic than that of Dawkins' memes borne by the stream of proceedings driven by genes.

The key idea, then, is that we humans tend to move increasingly into the 'virtual' spaces from the 'natural' one. It is not that we would be free to devise the 'virtual' spaces deliberately. One thing that prevents us from doing so is that the 'virtual' worlds cannot escape some embodiment in the sense of 'supervening' on the natural, physical space and having to respect all its possibilities and limitations fully. Another thing is that even the constitution of the 'virtual' worlds within these limits is not a matter of human will, but rather is 'led by an invisible hand'.

The Space of Meaningfulness

One of the crucial 'virtual spaces' opened for us thanks to the rules that we embedded in the foundations of our human societies is what I would like to call *the space of meaningfulness* – the space constituted by the rules of our language. It is the space which provides for the possibility of meaningful talk.

The ability to produce, protract and consume *meanings* was traditionally considered to be a characteristic feature of us humans. Meanings were usually also thought of as inseparably connected with the peculiar stuff of which our minds are made (and this was taken to be the explanation of the fact that they are an exclusively human affair). Hence the task of explaining language was seen as that of revealing meaning (typically a chunk of mental stuff) and explaining the way they are linked to expressions. Hence we must *first* explain meaning and *thereby* explain language, and *only then* can we possibly explain our linguistic practices.

However, during the last century the situation has changed rapidly.[3] A continuously increasing number of philosophers have tried to reverse this explanatory strategy: they have tried to account for our linguistic practices directly, leaving the concept of meaning at most the role of a – more or less useful – expedient of such an explanation (and if it turns out to be totally useless, the worse for it). This was the strategy of the later Wittgenstein urging us to see language as kind of toolbox, of Quine and Davidson with

their stories about radical translation or interpretation, and indeed of Sellars and Brandom, who have assimilated language to a rule-governed game.

The Sellarsian explanation of why we tend to see meaningfulness in terms of something glued to the expression might be that we tend to see expressions treated according to rules as acquiring *roles*; which then may get hypostasized and come to look like things.[4] Quine (1960) speaks about *analogical synthesis* here: we recognize the ways in which words function within some basic sentences and then extrapolate the ways so that we can assemble functionings of new sentences composed from known words.

But here again it might be useful to turn our attention to the *evolution* of language. Krebs and Dawkins (1984) conjectured that language as we know it came into being as a 'conspiratorial whispering'. Signals which, according to Dawkins and Krebs, originally evolved from the tendencies of organisms to predict the behavior of other organisms and from the counter-tendencies of organisms to exploit the fact that their behavior is being predicted for the purposes of manipulation of other organisms may further develop in two opposite directions. In cases where such manipulation harms the manipulated organism, the signals tend to require an increasing investment of energy till they become so costly that they fade away; whereas in cases when they are useful even for the manipulated, the energy invested may continually decrease and the manipulative behaviuor is reduced to mere 'symbols'. What makes the whole difference is the distinction between the 'competitive' and the 'cooperative' environment.

Hence cooperation, again. However, now the relation between a rule and cooperation is not as straightforward as in the Prisoner's Dilemma cases. Now we do not see following a rule as one side of a coin, the other side of which is cooperation; we rather see them as establishing a 'virtual world' which provides for a 'virtual' – or symbolic – signalling. As Knight (2008) puts it, whereas 'each animal can make a difference only physically, only with its body – with signals inseparable from the body', 'a human linguistic utterance – a 'speech act' – is an intervention in a different kind of reality. … A speech act, like a move in a game of 'let's pretend', is internal to reality of this kind.'[5]

On the face of it, the resulting claim sounds almost trivial: just as the rules of chess allow us to make pieces of wood into *bishops*, *rooks* and *queens* and play chess, the rules of language allow us to make various kinds of shrieks into contentful expressions and play our language games. But under this seeming triviality lurks a fantastically complex construction of rules: they are erected as barriers we bounce off as we bounce off the limits of our physical worlds (spelled out by the laws of nature). They interlock in multifaceted ways to open up virtual spaces where we can wield our freedom. They let us pass the rules and hence the spaces from generation to generation, so that they become not frail and transient, but solid and enduring. They let us enjoy the enigmatic forces of 'the virtual' without requiring us to devastate our bodies with drugs.

Conclusion

There are many suggestions aboout what makes us humans special: *soul*, *mind*, *language*, *culture*, *reason* …. In this paper I have indicated that we may characterize man as a normative being. Not that this proposal by itself would be original – of course it goes back at least to Kant; and recently a persuasive case for it has, in effect, been made by Brandom. However, I have tried to show that if we accept the analyses of the concept of *rule* put forward by Sellars, we can embed this characterization in the evolutionary stories of how we humans have become what we are.

I have tried to indicate that the crucial break which enabled man to live not only within the realm of nature, conforming to its laws, but also to enter the realm of freedom, where one can obey rules (while being free to *dis*obey them), has to do with the emergence of a behavioural 'meta-patterns, amounting to what Sellars calls an *ought-to-be* and making people comprehend and endorse patterns that they are taught.

Academy of Sciences of the Czech Republic and University of Hradec Králové, Czech Republic

Notes

1 American readers should not be confused by the fact that *football* is what they call *soccer*.
2 Von Wright (1963) calls them *directives*, whereas Raz (1999) speaks about *technical norms*.
3 I discuss this development elsewhere (see Peregrin, forthcoming).
4 See Peregrin, 2006.
5 See also Noble, 2000.

References

Axelrod, R. (1984) *The Evolution of Cooperation,* New York: Basic Books.
—— (1986) 'An Evolutionary Approach to Norms', *American Political Science Review* 80: 1095–1111.
Block, N. (1995) 'The Mind as the Software of the Brain', in D. Osherson, L. Gleitman, S. Kosslyn, E. Smith and S. Sternberg (eds) *An Invitation to Cognitive Science,* Cambridge, Mass.: MIT Press.
Dawkins, R. (1989) *The Selfish Gene,* Oxford: Oxford University Press.
Dennett, D. (2008) 'The Evolution of Why', in B. Weiss and J. Wanderer (eds) *Reading Brandom,* London: Routledge.
Fehr, E. and Gächter, S. (2002) 'Altruistic Punishment in Humans', *Nature* 415: 137–40.
Heckathorn, D. D. (1989) 'Collective Action and the Second-Order Free-Rider Problem', *Rationality and Society* 1: 78–100.

Knight, C. (2008) 'Language Co-evolved with the Rule of Law', *Mind and Society* 7: 109–28.

Krebs, J. R. and Dawkins, R. (1984) 'Animal Signals: Mind-Reading and Manipulation', in J. R. Krebs and N. B. Davies (eds) *Behavioural Ecology: An Evolutionary Approach,* Oxford: Blackwell, pp. 380–402.

Lehmann, L. and Keller, L. (2006) 'The Evolution of Cooperation and Altruism – A General Framework and a Classification of Models', *Journal of Evolutionary Biology* 19: 1365–76.

Maynard Smith, J. (1982) *Evolution and the Theory of Games,* Cambridge: Cambridge University Press.

Millikan, R. G. (2004) *Varieties of Meaning,* Cambridge, Mass.: MIT Press.

—— (2005) 'The Father, the Son and the Daughter: Sellars, Brandom, and Millikan', *Pragmatics and Cognition* 13: 59–72.

Noble, J. (2000) 'Co-operation, Competition and the Evolution of Pre-linguistic Communication', in C. Knight, J. R. Hurford and M. Studdert-Kennedy (eds) *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form,* Cambridge: Cambridge University Press.

Peregrin, J. (2006) 'Developing Sellars' Semantic Legacy: Meaning as a Role', in P. Wolf and M. N. Lance (eds) *The Self-Correcting Enterprise,* Amsterdam: Rodopi, pp. 257–74.

—— Forthcoming. 'Semantics without meaning?', in R. Schantz (ed.) Prospects of Meaning, Berlin: de Gruyter.

Poundstone, W. (1992) *Prisoner's Dilemma,* New York: Doubleday.

Quine, W.V.O. (1960) *Word and Object,* Cambridge, Mass.: MIT Press.

Raz, J. (1999) *Practical Reason and Norms,* Oxford: Oxford University Press.

Sellars, W. (1949) 'Language, Rules and Behavior', in S. Hook (ed.) *John Dewey: Philosopher of Science and Freedom,* New York: Dial Press, pp. 289–315.

—— (1954) 'Some Reflections on Language Games', *Philosophy of Science* 21: 204–28.

—— (1969) 'Language as Thought and as Communication', *Philosophy and Phenomenological Research* 29: 506–27.

Trivers, R. L. (1971) 'The Evolution of Reciprocal Altruism', *Quarterly Review of Biology* 46: 35–57.

von Wright (1963) *Norm and Action,* New York: Humanities Press.

Wittgenstein, L. (1969) *Philosophische Grammatik,* Frankfurt: Suhrkamp; English translation *Philosophical Grammar,* Oxford: Blackwell, 1974.

Woodcock, S. and Heath, J. (2002) 'The Robustness of Altruism as an Evolutionary Strategy', *Biology and Philosophy* 17: 567–90.