

Rules as the Impetus of Cultural Evolution

Jaroslav Peregrin

Topoi

An International Review of Philosophy

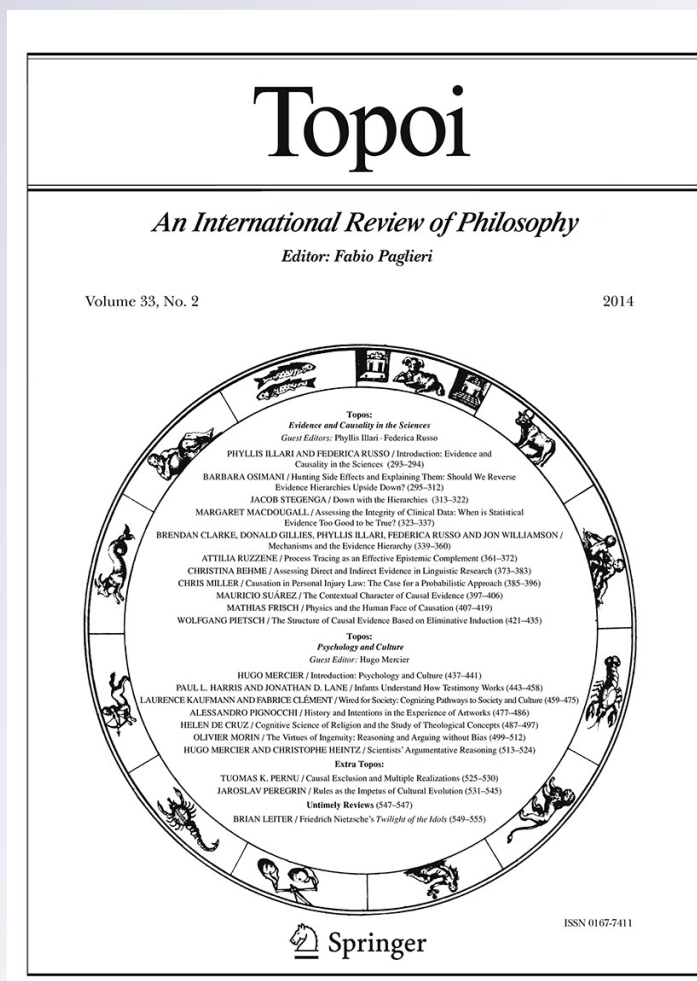
ISSN 0167-7411

Volume 33

Number 2

Topoi (2014) 33:531–545

DOI 10.1007/s11245-013-9219-2



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Rules as the Impetus of Cultural Evolution

Jaroslav Peregrin

Published online: 28 November 2013
© Springer Science+Business Media Dordrecht 2013

Abstract In this paper I put forward a thesis regarding the anatomy of “cultural evolution”, in particular the way the “cultural” transmission of behavioral patterns came to piggyback, through us humans, on the transmission effected by genetic evolution. I claim that what grounds and supports this new kind of transmission is a complex behavioral “meta-pattern” that makes it possible to grasp a pattern as something that “ought to be”, i.e. that transforms the pattern into what we can call a *rule*. (Here I draw especially on the philosophical insights of Wilfrid Sellars.) In this way I interlink empirical research done in evolution theory with some more speculative philosophical theories, thus shedding new light on the former and adding an empirical footing to the latter.

Keywords Rule · Evolution · Culture · Cooperation · Language

1 Cultural Evolution?

Since Darwin we have been able to explain the fact that organisms subtly adapt to their environment, without recourse to teleology. And since Konrad Lorenz, if not earlier, we have realized that the behavioral patterns displayed by organisms can be seen as on a par with their organs; hence we can view them likewise as products of

natural selection. Nevertheless, we know that, at least for us humans, this cannot be the whole story about behavioral patterns. It is obvious that in addition to the Darwinian evolution, such patterns have another mode of transmission across human societies: they are transmitted culturally, being demonstrated, taught, learned, imitated etc. This kind of handing over of behavioral patterns from generation to generation is perhaps less reliable, and generates patterns that are less robust, but it is incomparably quicker than the genetic one. And it seems that it is to this acceleration of the spreading of patterns that we humans owe most for what we are.

The fact that Darwinian evolution is too slow to account for what has been happening to mankind over recent millennia has, of course, been noted by many theoreticians; and the solution is also clear: the Darwinian, “hardware” evolution is piggybacked by a swifter, “software” kind of process—as already Wilson (1978, p. 78) noted, “cultural evolution is Lamarckian and very fast, whereas biological evolution is Darwinian and usually very slow”.¹ How is this cultural spreading of patterns possible and what are the mechanisms behind it?

¹ Tomasello (1999, pp. 2–4) exposes the reasons why we need to take cultural transmission of behavioral patterns seriously in the following way: “The fact is, there simply has not been enough time for normal processes of biological evolution involving genetic variation and natural selection to have created, one by one, each of the cognitive skills necessary for modern humans to invent and maintain complex tool-use industries and technologies, complex forms of symbolic communication and representation, and complex social organizations and institutions... There is only one possible solution to this puzzle. That is, there is only one known biological mechanism that could bring about these kinds of changes in behavior and cognition in so short a time whether that time be thought of as 6 million, 2 million, or one-quarter of a million years. This biological mechanism is social or cultural transmission, which works on time scales many orders of magnitude faster than those of organic evolution.”

J. Peregrin (✉)
Institute of Philosophy, Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Prague 1, Czech Republic
e-mail: jarda.peregrin@gmail.com

J. Peregrin
Faculty of Philosophy, University of Hradec Králové, Hradec Králové, Czech Republic

Being wired into human brains (or, for that matter, into brains of other animals) by natural selection and thus being projected into succeeding generations is not the only way a (lasting) behavioral pattern may exist. In principle, an individual might get canalized also by being rewarded when behaving in accordance with the pattern and penalized when not. But why would the individual's peers do this rewarding and penalizing? Perhaps *they* have this "normative" behavior wired in their brains by natural selection. Evolution theorists have introduced the concept of *altruistic punishment* to refer to something of this very kind. (More about this later.)

But this seems strange. Why would evolution enforce a pattern in such a detoured way, producing its "enforcers" forcing it upon "enforcees" instead of making the enforcees display it right away? And would it be possible at all? (As evolution theorists would insist, such an unconditionally altruistic individual would be hard-pressed to survive.) And even if this were possible, 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 the enforcers of the patterns are capable, not only of making the enforcees display it, but also of making the enforcees make others display it—hence they not only become adherents of the pattern, but also the pattern's enforcers. If this were the case, the pattern could be promulgated purely "culturally"; it would not need any wired-in support. In this way, the promulgation of behavioral patterns standardly effected by evolution would bear another level of such promulgation, piggybacking on it but going its own way.

My proposal, which I will back up in this paper, is, firstly, that the complex behavioral "meta-pattern" that makes us take a pattern as something that "ought to be", is what grounds and supports the cultural transmission of those behavioral patterns which are idiosyncratic to the human condition. Secondly, and interestingly, this is precisely what some philosophers have called a *rule*. Drawing upon these observations, we can interlink empirical research done in evolution theory with some more speculative philosophical theories, thereby shedding new light on the former and adding an empirical footing to the latter.

2 The Emergence of Rules

Most of the researchers addressing the phenomenon of human culture within the context of evolution see the root of our difference from other animal species in our enhanced ability to imitate our conspecifics. It is this ability, it is often argued, that lays the foundation of our ability to learn, to establish swift paragenetic channels of

passing skills, aptitudes and ultimately theories from generation to generation, and hence to create cumulative culture (Dawkins 1976; Donald 1991; Whiten and Cusance 1996; Meltzoff 1996; Tomasello 1999; Richerson and Boyd 2005). As Boyd and Richerson (2008, p. 306) put it: "cultural evolution leverages individual decision making by allowing individuals to acquire complex codes for behavior, mainly by the relatively cheap process of imitation".²

What I think is most conspicuously lacking from the picture of individuals imitating their more experienced or more successful conspecifics is assigning any prominent role to something like coercion, compulsion, forcing etc. Coercion is, as far as I can see, an inherent and crucial component of any enculturation and hence of the transmission and accumulation of culture. Even though we Westerners are proud that we no longer literally *hammer* morals (and other rules) into other's head, sanctions, punishment and compulsion still play a significant role in making neophytes comply with our rules.³

To be sure, the most primitive version of coercion, based on brute force, is not something that would set us humans apart from other animals and that could be seen as underlying our culture. This form of coercion amounts to the fact that the most basic attitude of an animal to its environment straightforwardly reflects its needs: it amounts to the exploiting of desired environmental resources and the elimination of hindering obstructions. There is also little doubt that where the environment consists not only of inanimate things, but also of other kinds of animals or other individuals of the same kind, the situation is similar; the primordial attitude is indiscriminately to utilize and adjust it in the light of one's needs. But the tendency to force others to fit one's needs by brute force is not what interests us here.

What is more interesting for us are various "cooperative" forms of coercion, i.e. where the coercion is in some or another sense "accepted" by the coerced party. A prominent case in point is teaching, where the teacher, in the case interesting for us here, forces the pupil into a behavioral pattern to which the pupil is not generally resistant (though she may be resistant to many of its concrete demands; hence the coercion). This specific kind of

² Of course, not everybody claims that the ability to imitate is *sufficient* for the emergence of culture. Candidate abilities which have been proposed for features that must be added include *shared intentionality* (Tomasello et al. 2005), *language* (Bickerton 2005), *recursive syntax* (Hauser et al. 2002), *pedagogy* (Gergely and Csibra 2006), *categorizing behavior* (Castro and Toro 2004) etc. We will be particularly interested in the last of these proposals.

³ Within twentieth century philosophy, Michel Foucault became notorious for stressing—perhaps excessively—this aspect of acculturation (see esp. Foucault 1975).

“cooperative coercion” is already largely specific to us humans.⁴ However it seems to be a vital addition to the spreading of patterns through mere imitation, for it corrects the errors of imitation and minimalizes learning costs by evading many of the perils of trial-and-error learning. (This was stressed by Castro and Toro 2004, and by Gergely and Csibra 2006.⁵) Also, we have ways to explain the evolutionary *rationale* of our investment into teaching.⁶

Consider a (“first-order”) behavioral pattern, like always walking on the right side of the road, avoiding looking at a chief, or greeting everybody in the morning. Call the behavior used to enforce this behavioral pattern the *associated second-order pattern*. Hence the second-order behavioral patterns associated with the above first-order patterns would be forcing others to always walk on the right side of the road (by way of rewards and punishments), forcing them not to look at the chief, or forcing them to greet everybody in the morning, respectively. Obviously, a first-order pattern and its associated second-order pattern are two different patterns that might be acquired, displayed, enforced etc. independently of each other. (In principle I can be taught to (1) avoid looking at a chief and to (2) teach my children not to look at the chief as two independent lessons.) However, it would seem more economical to have a general propensity for teaching one’s children what one is taught by one’s parents. If this is the case, then being taught something may involve being taught to teach the something.

The importance of this propensity is even more apparent when we move from teaching to the more general version of enculturation effected by social norms. Castro et al. (2010, pp. 352–353) describe the situation as follows:

Here we propose that, during ontogeny, the assessor communication between parents and offspring is extended by other evaluative interactions where the approval or disapproval of behavior is provided by other unrelated individuals. Throughout their lifespan, a person establishes a social reference group

with individuals that interact closely during a particular stage of life (parents, partner, friends, and colleagues). Our thesis maintains that the individuals in the social reference group are neither neutral nor passive towards the behavior of a person in the group. Rather, they evaluate and demonstrate approval or disapproval, even if the behavior in question does not affect them directly. Chimpanzees may classify other individuals’ behavior as favorable or unfavorable with respect to themselves, and may act accordingly, but the ability to approve or disapprove of other individuals’ learned behavior seems completely absent in primates (...).

The difference is that while in the case of pedagogy we have fixed roles (of teacher as the authority and pupil as subordinated to the authority) which are replicated from generation to generation (pupils becoming teachers), in the case of (“cooperative”) social norms there is, usually, no explicit authority. Anybody is subordinated to the authority generated by everybody. Hence one cannot assume the role of being subordinated to norms (“pupil”) and at some later point switch to the role of the authority enforcing the norms (“teacher”). Thus, here the merging of *being squeezed into a pattern* and *being made to squeeze others into the pattern* becomes imperative.

What I am suggesting as the crucial specific evolutionary innovation of us humans is this very behavioral meta-pattern which causes us to adopt also the associated second-order pattern when being forced into a given first-order pattern—in other words, which makes us take what we are forced to do also as something that generally *should be* done. We humans have not only developed a capacity to approve or disapprove of others’ (especially our offsprings’) behavior and a sensitivity to expressions of approval and disapproval by others (especially our parents), but also a capacity to accept certain expressions of approval and disapproval by others both as guides for our own behavior, and, simultaneously, as guides for our approving/disapproving of others.

Tennie et al. (*ibid.*, p. 2412) write:

This normative dimension to human cultural traditions may be seen again as deriving from teaching, in this case adults teaching children how to and how not to do things. But this is not totally accurate, as it was not the case in either the Rakoczy et al. (2008) or the Kelemen (1999) study that children observed adults making normative judgements about the right and wrong actions involved. So they were not copying adults teaching others what not to do; the children spontaneously jumped to the conclusion that the way the adults were showing them how to do it was the right way, to which everyone should conform. The

⁴ Tennie et al. (2009, p. 2411) write: “Teaching is present in all human societies we know of (Kruger and Tomasello 1996), and it is clearly not an everyday activity among chimpanzees or other non-human primates—though something in this direction may occur occasionally (...).” See also Warneken and Tomasello (2009).

⁵ Gergely and Csibra (*ibid.*, p. 241) argue that “pedagogy” is “a primary species-specific cognitive adaptation to ensure fast, efficient, and relevance-proof learning of cultural knowledge in humans under conditions of cognitive opacity of cultural forms”.

⁶ See esp. Simon (1990, p. 1665): “Because docility—receptivity to social influence—contributes greatly to fitness in the human species, it will be positively selected. As a consequence, society can impose a ‘tax’ on the gross benefits gained by individuals from docility by inducing docile individuals to engage in altruistic behaviors. Limits on rationality in the face of environmental complexity prevent the individual from avoiding this ‘tax’”. Cf. also Castro and Toro (*ibid.*).

evolutionary source of this normative dimension to human activities is not immediately clear, although it is surely bound up in general, one would think, with group identity and conformist transmission (this is how we, the members of this group, do things—even if others do them differently)—as enforced by punishment, including third-party punishment.

I will call the tendencies to generally approve or disapprove of others' behavior, borrowing from Brandom (1994), *normative attitudes*. They are what explicit pedagogical guidance becomes when being generalized, shifted to the level of social norms, and becoming much less explicit; and they form the kind of “social friction” one is trained to become sensitive to during one's education and enculturation. They are, I am convinced, the primary sources of any normativity.

What is important is that these attitudes need not be understood as a matter of propositional attitudes. In many cases they are merely *practical* attitudes which we characterize in terms of characterizing the corresponding behavioral pattern, without presupposing anything about the cognitive capacities of the ones displaying the attitude. (This is because the ability to follow rules, according to our approach, is more fundamental than propositional thinking and hence cannot be explained in its terms.)⁷

Note also that while teaching usually precedes direct contact with social rules in the course of ontogeny, from the viewpoint of phylogeny, social rules may be more fundamental, both because teaching, at least in its systematic form, presupposes at least some partly social framework, and because what is taught during enculturation consists mostly of (correlates of) social rules. Hence it would be unlikely that the ability to create, maintain, obey etc. social rules is just a more developed form of docility. Thus, the meta-pattern taking us from regularities to rules is crucial especially with respect to these rules' bearings on social cooperation.

What I suggest is that when one acquires this complex behavioral meta-pattern (taking us from learning certain first-order patterns to acquiring also the associated second-order patterns, from becoming forced into a behavioral pattern to becoming its enforcer) one becomes a rudimentary rule-follower. Thus we can say that it is this general kind of meta-pattern that makes us into rudimentary normative creatures. It is, it seems, a meta-pattern that is mostly beyond the ken of other kinds of animals; and it is this meta-pattern that may be responsible for the fact that our kind found a swifter way of spreading behavioral patterns than that offered by genetics—a way that has made us so radically different from our animal cousins. And this,

⁷ We can speculate that the normative attitudes may be based on something akin to Gendler's (2008a, b) *aliefs*.

I think, provides a very interesting, new reading of one of the traditional philosophical characterizations of man, viz. a *normative being*. In what follows, we will first survey a proposal of Wilfrid Sellars concerning rules, and then return to discuss the consequences of this alleged “normative turn” of our animal predecessors.

3 Sellars' “Pattern-Governed Behavior”

In philosophy, there have always been plenty of discussions about what it is that makes us humans special. Obvious answers may be *mind*, *language* or *culture*; but these answers are of limited use, for what is truly interesting is what *underlies* these phenomena: what was it that made *Homo sapiens*, in contrast to other animals, able to develop them in the first place? The answer to which we are led by the considerations put forward in the present paper is that it is *rules*—and hence that we humans are first and foremost *normative* beings. This answer chimes with views that go back to Kant, and in the twentieth century they were vigorously revived especially by Wilfrid Sellars. As Sellars (1949, p. 298) puts it: “When God created Adam, he whispered in his ear ‘In all contexts of action you will recognize rules, if only the rule to grope for rules to recognize. When you cease to recognize rules, you will walk on four feet.’”

This might seem *prima facie* implausible. Do we not need language to formulate rules? And do we not need mind to have language? We may imagine that *first* we humans came to have mind (due to evolution equipping us with bigger brains), *then* we came to have language (for making one's thoughts public turned out to be useful) and only then we came to have rules (for after we came to be able to talk to each other, we might find it advantageous to regulate our social relationships). But the Sellarsian view is very different: rules are what underlie language, and language is what underlies mind—at least the kind of mind we humans have. This is not to say that we must *first* be fully-fledged rule-followers, *then* fully-fledged language-users, and *only then* can we have our kind of mind; of course some kind of bootstrapping must be in play. First we have some rudimentary rules, then some rudimentary language and rudimentary human-kind mind, which then helps us have more fully-fledged rules etc.

The point of this Sellarsian answer is that though language, of course, is based on some biological predispositions, it is, in essence a social institution and as such cannot escape the normativity (rule-baseness) of any social institution; and that the same is true about our human way of thinking. This is what makes the situation, according to Sellars, so complex: if language could be exempted from normativity, rules could be always already articulated in language; but as this is not the case, we have to count with

some “inarticulate” rules (in pain of infinite regress) and we must explain what is the nature of such rules.⁸

Sellars' theory drew upon the idea that there is an important parallel between language and games, in that both are *essentially rule-governed enterprises*. What does this mean? Both language and games like chess or football are *constituted* by rules. There is no scoring a goal aside of the framework of the rules of football. Whatever spherical thing you move through whatever gate-shaped construction, you are not scoring a goal if this is not a part of a football match, which in turns does not happen if the rules of football are not in force. (And notice that this is not restricted to the “officially rule-governed” football, i.e. to matches with referees, spectators and TV reporters; even football as played by kids in backyards must have *some*—though perhaps vague and never quite spelled out—rules.) And whichever sounds you emit or whatever marks you make, you are not saying *that it rains* aside of the framework of a language with *its* rules.

Moreover, the rules of both football and language have the character of *constraints*. They do not usually *prescribe* us what exactly to do, they tell us what is prohibited. And by spelling out what is prohibited they erect certain limits, and by erecting the limits they set up an internal space. In the case of football, this is a space which can offer one—either directly as a player or indirectly as a spectator—exciting and breathtaking adventures; adventures of a kind simply unavailable outside of the space. In the case of language, the adventures are perhaps not so directly exciting, but perhaps even more breathtaking. They are the adventures of meaningful talk. Hence, we can say, the rules of language open up the space of meaningfulness.

The toughest problem for this view of language lies in making sense of the notion of an *implicit rule*—a rule that is not introduced as an explicit prescription, but rather emerges spontaneously. Sellars' idea was that linguistic behavior is a species of the general kind of behavior that, though not governed by the *conception* of any *explicit* articulation of rules, is, in a sense, governed by rules over and above being simply regular in the way complying with the rules. Thus, this kind of behavior is more than a behavior that merely displays regularities, but less than a conscious and purposeful rule following.

To envisage how Sellars (1949, p. 297) himself sees the matter, let me quote him at length:

⁸ The case for the rules of language not being able to be generally explicit was made, most famously, by Wittgenstein (1953). His point was that an explicit rule, a rule in the form of a “symbolic” (linguistic) object, must be *understood*, i.e. we must be able to interpret it, and if we agree that making something into a symbol, i.e. granting it a meaning, is a matter of rules, a successful interpretation must be a matter of us using the right rules—hence coming to understand a rule would presuppose understanding other rules.

Clearly the type of activity that is rule-regulated is of a higher level than that which is produced by simple animal learning procedure. One way of bringing this out is to say that most if not all animal behavior is tied to the environment in a way in which much characteristically human behavior is not. Certainly we learn habits of response to our environment in a way which is essentially identical with that in which the dog learns to sit up when I snap my fingers. And certainly these learned habits of response—though modifiable by rule-regulated symbol activity—remain the basic tie between all the complex rule-regulated behavior which is the human mind in action, and the environment in which the individual lives and acts. Yet above the foundation of man's learned responses to environmental stimuli—let us call this his tied behavior—there towers a superstructure of more or less developed systems of rule-regulated symbol activity which constitutes man's intellectual vision.... Such symbol activity may be well characterized as free—by which, of course, I do not mean *uncaused*—in contrast to the behavior that is learned as a dog learns to sit up, or a white rat to run a maze. On the other hand, a structure of rule-regulated symbol activity, which as such is free, constitutes a man's understanding of *this* world, the world in which he lives, its history and future, the laws according to which it operates, by meshing it with his tied behavior, his learned habits of response to his environment. To say that man is a rational animal, is to say that man is a creature not of *habits*, but of *rules*.

The problem of understanding the role of rules within human linguistic conduct, then, can be portrayed as that of steering between the *Skylla* of *regulism*, which is a matter of claiming that a rule is by its nature *explicit* (we have already seen that this leads to a vicious circle), and the *Charybdis* of *regularism*, claiming that rule-governed behavior is nothing more than regular behavior (which would erase any difference between a stone's following the law of gravitation by falling and a person's following the rule of traffic by stopping at a red light).⁹ Hence, Sellars (1954) suggests that our language games are a matter of a specific kind of behavior which qualifies neither as “merely conforming to rules”, nor as fully-fledged “rule obeying”. He calls this kind of behavior *pattern governed*: “an organism may come to play a language game—that is to move from position to position in a system of moves and positions and to do it ‘because of the system’ without

⁹ See Brandom (1994, Chapter I) for a thorough discussion of this issue.

having to *obey rules* and hence without having to be playing a *metalinguage game*" (*ibid.*, 209).

This is related to Sellars' distinction between what he calls *ought-to-do* and *ought-to-be* rules: whereas the former ones are straightforward directives for action, the latter yield such directives only indirectly: if something ought to be thus and so, we ought to do what can contribute to its being thus and so. A special case of an *ought-to-be* is a behavioral pattern. Tutors who endorse this *ought-to-be* derive the obvious *ought-to-do* rule: they educate their tutees so as to bring about their displaying the pattern. But if this were the whole matter, the consequences of the tuition would be limited: the tutees would come to comply with the *ought-to-be* pattern, but the pattern would not outlast their physical existence. What is needed to achieve the perpetuation is that the tutees not only get the pattern, but get the corresponding *ought-to-be* as a *rule*, from which they derive their *ought-to-do* consequences, which makes them educate their tutees. And so on *ad infinitum*, or at least till the human chain is severed.

This institutes the very kind of circle that, as we indicated above, is *reproductive* in the sense that it provides for the "cultural" spreading of behavioral patterns piggybacking on the "natural" spreading effected by evolution. The relevant patterns are forced upon us not (directly) by natural selection, but by the ongoing 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).¹⁰

4 Rules and Cooperation

The meta-pattern I deem responsible for rule following and for the emergence of Sellarsian *ought-to-be*'s (and for the turn in humankind's history which has largely freed our developments' trajectory from the direct control of our genes) is, we saw, our ability to turn regularities into rules. We can accept patterns into which we are forced as not only something to which we must submit ourselves (or perhaps against which we can try to revolt), but as something which we *should* accept. An obvious question, then, is where do the regularities which are thus turned into rules come from—what is the source of the behavioral patterns which we start to see as the *ought-to-be*'s?

A behavioral pattern may result from sheer coercion. (If, for example, another individual beats me near to death whenever I do not share my spoils with him, my behavior is

¹⁰ See Peregrin (2010) for a more detailed discussion of Sellars' standpoint.

	C	N		C	<u>N</u>		C	N	
C	2,2	0,1		C	2,2	0,3	C	1,1	0,0
N	1,0	1,1		<u>N</u>	3,0	1,1	N	0,0	1,1
	Stag Hunt			Prisoner's Dilemma			Pure Coordination		

Fig. 1 The three most basic games relevant for the study of cooperation. Nash's equilibria are printed in boldface. Pareto efficient strategic profiles are printed on the grey background. Dominant strategies are underscored

likely to acquire a certain shape.) But for obvious reasons, such a pattern is an unlikely candidate for the one which I will perceive as an *ought-to-be* in the Sellarsian sense. It would seem that a much more plausible aspirant would be a pattern arising out of what we called *cooperative coercion*. Teaching, certainly, would be an important source; but here again we face the question of the source of the patterns we are forced into during education. Certainly in part these may concern the skills instrumental to direct coping with nature (and hence are shaped, as it were, by its resistance), but the part more interesting for us here lies in the teaching effecting the implementation of social rules. Where do the patterns which carry these rules (i.e. the regularities that are being turned into the rules) come from? And the only available answer seems to be that they must emerge, spontaneously, from some coordination of our activities.

The emergence of cooperation is usually associated with two kinds of game-theoretical models, accounting for the concepts of *coordination* and *reciprocation* (or, somewhat misleadingly, *altruism*), respectively. The first of them is, in the simplest case, associated with the Stag Hunt game,¹¹ in which cooperation naturally flourishes because it helps both parties to a better outcome; while the second is associated with the Prisoner's Dilemma game,¹² which pictures the origin of cooperation as somewhat mysterious, because it conditions the improvement of the outcome by future reciprocation, making the rejection of cooperation seem more profitable.¹³

In the simpler case of coordination, however, the situation may be such that it is difficult to find the profitable equilibrium, as in Pure Coordination game. In this situation (unlike in the simple Stag Hunt case) we have multiple strategies, the payoff for each of which equally depends on how many other players choose the same strategy. If all or

¹¹ See Skyrms (2004).

¹² See Pounstone (1992). In evolution game theory we often encounter the closely related Hawk and Dove game—see Maynard Smith (1982).

¹³ For accounts of human sociality with greater stress on the first concept see, e.g., Skyrms (1996) or Binmore (2005); for those focusing on the second, see, e.g. Axelrod (1984) or Bowles and Gintis (2011).

almost all users fall upon the same strategy, it is this strategy that becomes profitable (and we can say that it is such purely conventionally, for its profitability is not a matter of the rules of the game, but rather of the fact that it was raised to the status by the mere resonance of the strategies of the players) (Fig. 1).

Difficulties arise when the convention cannot be explicit, i.e. when the players cannot agree on the strategy which is to become the profitable equilibrium. How, then, should I choose my strategy? In this case, the convention must build on some previous foundations, on the facts that some of the strategies have become prominent in terms of a habit or have become salient in some other way.¹⁴ To make such a habitual choice into a useful convention is to treat it as something that ‘should be’—something anybody should follow for everybody’s sake.¹⁵ Hence again, we would require the ability to treat a certain extant behavioral pattern as something that *ought to be*.

The upshot, then, is that whatever the convention is, cooperation is sustained by the fact that it is treated as something that should be followed. Something that may have evolved as a mere habit will become a framework of cooperation once it “goes normative”: a behavioral regularity is turned into a rule, into something that should be done (and should be seen to be done).¹⁶ Thus, it is a pattern emerging as a spontaneous solution of as a coordination problem that our meta-patterns comes to be applied to.

The Stag Hunt and the Coordination Game situations, are both situations in which cooperation comes about smoothly, for one does not profit from non-cooperation. Hence I just choose a strategy which is optimal for me; the only problem might be to identify it, viz. to align myself with other participants. Many authors, however, argue that “true” cooperation requires that I depart from the strategy which is optimal for myself as an individual and invest something into the society, and perhaps only much later be rewarded. Doing this is profitable only if everybody (or almost everybody) does the same; but unlike in the case of pure coordination, here one can profit from not doing what the others do, which makes it hard to explain why the participants would come to cooperate at all.

Most solutions to this situation proposed in the literature have presupposed that the parties of the Prisoner’s Dilemma are endowed with something more than merely one of the two available strategies, and hence that, ultimately, it is a more complicated game derived from the

simple Prisoner’s Dilemma that we should consider. Often it is the Iterated Prisoner’s Dilemma—the game consisting of an unlimited number of repetitions of the Prisoner’s Dilemma.¹⁷ The strategies of this game, then, offer a much wider space of possibilities, in which there exist equilibria of cooperating strategies.¹⁸

A vital extension of this approach to the Prisoner’s Dilemma is the inclusion of a social dimension, whereby a player has a degree of choice over whom to play with next—in this case a cooperative player may struggle to avoid non-cooperating peers.¹⁹ This can be seen as rewarding the cooperators by admitting them into an “exclusive sub-community” we form together with them,²⁰ while punishing the non-cooperators by ostracizing them. Variations on such models have been studied in the literature and shown to be able to support cooperation, hence overcoming the barrier of the Prisoner’s Dilemma.²¹ A special case of this problem, intensively studied, is how we could have come to agree on the principles of morality.²²

To sum up: social contact may bring about a need for coordination or reciprocation, which in turn produce certain shared behavioral patterns, certain general regularities of behavior, but their stabilization requires that these are turned into rules. The ensuing pattern must come to be treated as something that *should be*—hence must be backed up by the meta-pattern discussed in this paper. It is as the result of this that cooperators come to be rewarded, and non-cooperators punished, and this is what stabilizes cooperation and prevents it from being eroded by ignorance or free-riding.

5 Rules, Language and Communication

Let us now consider, in greater detail, the problem of the emergence of language as the most important component—and vehicle—of culture. We must start farther back than with creatures using signals—signals already have to exist within a kind of coordinative (if not rule-based)

¹⁷ See Le and Boyd (2007).

¹⁸ Axelrod (1984) was the first to show that in such settings strategies like tit-for-tat need not be doomed to extinction. More exact characterization is offered by the so called “folk theorem” (see, e.g., Binmore 2005).

¹⁹ See McKenzie Alexander (2007) or Spiekerman (2009).

²⁰ Tomasello et al. (2012, p. 673) speak about the psychological background of this process as a “new group-mindedness” that “creates cultural conventions, norms, and institutions (all characterized by collective intentionality), with knowledge of a specific set of these marking individuals as members of a particular cultural group”. See also McElrath et al. (2003).

²¹ See Woodcock and Heath (2002), Nowak (2006), or West et al. (2007).

²² See Joyce (2006) or McKenzie Alexander (2007).

¹⁴ See the elaborations of Lewis’ (1969) theory of convention by Cubit and Sugden (2003) or Sillari (2013).

¹⁵ See Tummolini et al. (2013).

¹⁶ Guala (2012) argues that the normative dimension is not *added* to the originally non-normative convention; that the Lewisian convention cannot but be normative from the very beginning.

framework, whereas we want to consider how such a framework might get established in the first place.

Hence let us return to the point where we left our ancestors acquiring the meta-pattern of taking something as an *ought-to-be*, i.e. not only submitting to a coercive force, but also, in one sweep, joining the enforcers. Such a situation would have led to the emergence of a certain order, of setting certain limits to what can be expected from others (and what can be feared from them). One of the important (though perhaps unexpected) outcomes of this process is that there arises the possibility of influencing the behavior of the members of such a proto-society, unlike the functioning of the rest of the environment, in a merely “symbolic” way (i.e. in a way not requiring a substantive investment of energy—instead of making it physically impossible for somebody to do something, I have the possibility of preventing him from doing it simply by expressing my disapproval.)

Though any account of the emergence of anything like signals has to remain speculative, I think that the account sketched by Krebs and Dawkins (1984) fits very well with our line of argument. Their story of the origins of language pictures it emerging out of the interplay of “mind-reading” and “manipulation”. The fitness of a creature from the viewpoint of natural selection obviously increases with its ability to predict the behavior of its environment; including its living environment, i.e. other creatures. Hence it is to be expected that evolution will produce creatures that will be able to “read the minds” of other creatures, i.e. predict what these creatures are about to do. (Of course, the term “mind reading” is used metaphorically—neither the reader, nor the creature being read must have anything that we would call a mind. The point is just that evolution molds the creature so that it behaves *as if* it can read the mind of another.)

Now, given that creatures compete, any enhancement of fitness tends to prompt a countermeasure. The countermeasure that emerges here is what Krebs and Dawkins call “manipulation”. A creature whose behavior is being predicted (whose “mind” is being “read”) would take advantage of this fact and would use the advantage to manipulate those attempting to read its mind. The predicting produces a situation in which some hints in the behavior of the predictee lead to a certain reactive behavior of the predictor, who is awaiting the given behavior of the predictee. Hence the predictee might “tease” the predictor into displaying this reactive behavior by displaying the relevant hints; and in this way it can manipulate it.

Escalation of these “arms races”, according to Krebs and Dawkins, may lead to two different outcomes, according to the nature of the emerging manipulation. If the manipulation does not accord with the interests of the manipulated, then of course the counter-countermeasure of the manipulated is to stop taking the hints seriously; as a

result of this, they keep working only with higher and higher energy investments until they become completely unusable. If, on the other hand, they do accord with its interests, the hints can gradually reduce to a total minimum, which Krebs and Dawkins call “conspirational whisper” (thus becoming what someone might want to call “symbolic”). And this, they claim, is the kind of signaling that constitutes a rudiment of language.

Note that in this simple situation the possibility of cheating or free-riding is limited. From the game-theoretical perspective (see Skyrms 2010), if the interests of the sender and the receiver coincide, the game is that of pure coordination, where the only problem is to zoom in on a common strategy, to achieve a conventional equilibrium. A cheater is in no better position than he who would “cheat” in the coordination game in which the participants have reached the convention of always walking on the right side of the street. On the other hand, if the interests of the sender and the receiver diverge, there is no equilibrium at all and hence nothing to deviate from by cheating.

However, as Skyrms (*ibid.*, p. 73) himself points out, we know that there *is* cheating in various signaling games and hence that “Any theory that says that deception is impossible is a non-starter.” Why is this? Well, once signals come to have established *roles* (“meanings”) in the signaling game, one may pretend to “mean” something without really “meaning” it (to some profitable effect). It is still true that *global* cheating is impossible (for this would simply rob signals of any roles), but local abuse is eminently possible. This is clearly illustrated by our use of our contemporary developed language: if everybody (or nearly everybody) always (or nearly always) violated its rules (for example, by asserting falsities and agreeing with them), the rules would probably soon disintegrate, but sinning against the rules now and then (telling occasional lies) is not only possible, but may be wide-spread.

What differentiates a simple signaling game in which cheating makes no sense, from a more developed enterprise where one can cheat (and will be tempted to)? We have indicated that it is the established roles the signals have. If each utterance of a signal contributes to the constitution of its role, then there is no misuse (but rather only a shifting of the role); it is only when the roles are held in place independently of individual utterances that I can use a signal “out of its role”. How can such a holding in place be effected?

One possibility is that there emerges a *regularity* and a habit. We come to use a signal habitually in the case, say, of danger; hence if it is used when no danger is in sight, it is *misused*. However, is it used incorrectly? And is there any pressure which would hold such “misuses” at bay? Maybe the one to misuse it just has a slightly different habit amounting to using it in some exceptional cases even in the cases when there is no danger? Maybe the habits of the

whole community have appeared to match only up to now; and from now on they will dramatically diverge?

Another possibility is that there is not only the regularity, but a *rule*—hence there would exist normative attitudes to the signals rendering them right or wrong. This framework shifts the burden of sustaining the roles completely to the normative attitudes, and makes space for using the signals out of their roles—to mistakes, misinformation or cheating.²³ This framework, then, requires elementary *trust*: “I will take your hint seriously because I *trust* you that you are serious and benevolent.” (Of course, this should not be read as attempting to reconstruct some inner monologue of the protagonist, for the protagonists are as yet pre-linguistic creatures; hence this is merely to metaphorically render the rationale put into the behavior by evolution.) However, once the framework of trust is grounded in *ought-to-be*'s, i.e. in the normative attitudes participants assume to the behavior of others (as well as to their own), there is an automatic ostracizing of the frustrators. It is this systematic chastising of those who cheat and praising of those who do not, that can lead to a truly stabilized form of signaling.

This problem largely overlaps with the discussion about the nature of conventions on which many philosophers claimed language is based: It seems clear that the fact that the word “dog”, in English, means what it does is nothing that would be implied by this very type of sound or inscription, it is we, speakers of English, who have associated it with this particular meaning. Hence it seems that before there can be a language, there must be the establishment of the conventions underlying it. But here a vicious circle looms: how can we establish a convention to bring language into being without already having a language?

We already saw that there is a way for conventions to emerge out of the spontaneous solution of coordination problems. Lewis (1969) was the first to investigate such implicit conventions that might underlie language in detail. Lewis' conclusion was that the kind of coordination problems that can be seen as yielding rudiments of language can be modeled in terms of certain “signaling games”.²⁴ And as we already saw, Lewis' followers then made it explicit that the convention involved in this cannot fail to engender a normative dimension, which is effected by our tendency to turn this kind of regularities into rules. This is our meta-pattern at work.

²³ To be sure, normative attitudes may also conceivably come to diverge, which would mean the disintegration of the corresponding rule. However, as long as they are in place, they constitute the roles which an individual move may fail to respect, and this, in some cases, may bring him some advantage.

²⁴ The notion was later elaborated by a number of authors; see especially Skyrms (2010).

Now as Knight (2008, p. 124), reflecting on Krebs and Dawkins, points out, while doing anything significant in the natural world requires the exertion of much energy, the world constituted by the rules of language provides for the possibility of achieving (a peculiar kind of) significance at virtually zero energetic cost:

Each animal can make a difference only physically, only with its body—with signals inseparable from the body.... By contrast, 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.... When human life became subject to the rule of law, participation in this kind of reality became possible for the first time. Because signals internal to this novel domain were no longer evolving in a Darwinian world, the familiar laws of signal evolution (...) no longer applied. Intrinsic reliability was no longer a requirement, allowing zero-cost signaling to emerge. Among other consequences, messages could now be encoded as digital shorthands. Thanks to this remarkable development, abstract principles such as recursion—formerly restricted to internal cognition—could now for the first time find expression in public language. The former paradoxes of language evolution correspondingly dissolved.

Let me add that to say that the fact that communication requires *trust* is to say that it requires *responsibility*: being a trustworthy speaker, i.e. a speaker whose sounds are to be taken as truly meaningful, rather than as mere babbling, is a status that is hard-won.²⁵

As a consequence, what I think is essential for a collection of signals to become what we call *language* is not only the breaking through of a certain “barrier of complexity”,²⁶ but also the emergence of what Sellars calls

²⁵ The importance of such statuses—under the headings of “prestige” or “reputation”—has been stressed by many authors (see, e.g., Frank 1988, or Henrich and Gil-White 2001). And to avoid misunderstanding, let me also stress that connecting the human kind of communication with trust and responsibility is not saying that lying, deceiving or hiding one's intentions would be impossible or marginal. But the framework which makes communication, and hence also deception, possible, cannot be itself based on deception—we saw that there is nothing to be gained from violating purely conventional norms.

²⁶ The “barrier of complexity” that must be broken through is not only a matter of the fact that a language worth its name must consist of a great number of signals, but also of the fact that the signals come to constitute a “structure”. Some of them become *incompatible* with other signals in force of being bound to situations which cannot co-occur (thus a signal indicating danger comes to be incompatible with a signal indicating the absence of any danger). Some of them become *inferable* from other signals (thus a signal indicating danger comes to be inferable from a signal indicating an approaching tiger).

“linguaging about languagins” (i.e. the second-order—linguistic—behavior associated with the first-order linguistic signals), i.e. reactions to the newly emerging signals, prominently including critical reactions. In this way, what emerges is a truly discursive realm—a relatively self-contained domain of interlocking linguistic actions. In this way, certain signals become tied to the situations where they are used not merely habitually, but rather *normatively*: they are something that is taken to be *correctly* displayed in such a situation (primarily in the sense that they should never be used in different kinds of situations, sometimes secondarily also in the sense that they should always be used in this kind of situation).

I think that this might be generalized: it is not only *linguistic* actions that may invite such “feedback loops”, in connection with which we, in retrospect, use terms like *trust*, *cooperation*, *sanctions* or *rewards*; for all kinds of social actions have a “symbolic” dimension insofar as having consequences not merely as physical actions in a physical world, but rather consequences drastically amplified by means of the institutional frameworks within which they take part.

6 Culture and Its Transmission

The idea that at some stage 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) became influential especially after Dawkins (1976) coined the suggestive term *meme*. The idea that the working of the fundamental replicators, genes, gives rise to memes as a different form of replicators and hence that evolution as if replicates itself on a higher, more abstract and swifter level, is no doubt seductive, but Dawkins himself did not offer much more than a gesture towards such a theory. Moreover, it seems that his straightforward assimilation of “memetics” to “genetics” seems to be too simplistic.²⁷

Like many other authors thereafter (cited at the start of this article), Dawkins invoked the notion of *imitation*. But as I have already indicated, though I think the ability to imitate no doubt contributes vitally to the ability to promulgate behavioral patterns “culturally”, by itself it is insufficient to account for its complexities. What makes us humans unique, what makes our antics, in contrast to those of other species, warrant 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.²⁸

²⁷ Cf. Gabora (2008).

²⁸ Dawkins tries to account for this in terms of *imperfections* of the way we copy memes—people, according to him, often do not quite

Hence imitation requires other essential ingredients to provide for the cultural promulgation of behavioral patterns relatively independent of the genetic one. The Dawkinsian account, and other accounts similarly based purely on imitation, makes us see the upgrading of ideas as a matter of their transfer from one individual to another. But in reality, 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, but it is to say that if we are to talk about “memes”, we need to see them as essentially distributed. They do not exist in individual humans, but in networks of human interaction within human societies.

What I think is again missing from the accounts based on imitation, is any reflection of the fact that not only our theories, but also our societies in general evolve largely via various forms of confrontation, ranging from friendly arguments to hostile conflicts. Concentrating on the milder forms of conflict, we can agree with Rochat (2006) that *Homo sapiens* is basically *Homo negotiatus*—any kind of human intercourse, including those that effect acculturation, quickly evolve into negotiations, that is, into a kind of interaction which, unlike imitation, is not characterized by the purely passive role of one of the parties.²⁹

Boyd et al. (2011) make this point very clear in their criticism of the so-called “cognitive niche hypothesis”: the thesis, as they construe it, based on the assumption that culture arises out of the fact that human development essentially accelerates when only a minority of individuals acquires new skills by means of a direct investigation of their natural environment and the rest of them imitates these foragers. The alternative Boyd et al. (*ibid.*, p. 10923) propose, in place of this, what they call the “cultural niche hypothesis”:

The cultural niche hypothesis and the cognitive niche hypothesis make sharply different predictions about how local adaptations are acquired and understood. The cognitive niche hypothesis posits that technologies are adaptive because improvisational intelligence

Footnote 28 continued

imitate one another, but do it only imperfectly. (Thus, he, 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 least it seems that *imperfection* is a very inadequate word to characterize the difference between mere imitation and the actual upgrading which is really going on.

²⁹ Rouse (2007), whose stress on normativity of social practices is congenial to the view advocated here, characterizes the complexity of the relevant mode of interaction in this way (p. 49): “One performance responds to another, for example, by correcting it, drawing inferences from it, translating it, rewarding or punishing its performer, trying to do the same thing in different circumstances, mimicking it, circumventing its effects, and so on.”

allows some individuals figure out how they work and why they are better than alternatives. These acquired understandings of the world are then shared, allowing others to acquire the same causal understanding without costly individual investigation. In contrast, we argue that cultural evolution operating over generations has gradually accumulated and recombined adaptive elements, eventually creating adaptive packages beyond the causal understanding of the individuals who use them. In some cases elements of causal understanding may be passed along, but this is not necessary. Often individuals will have no idea why certain elements are included in a design, nor any notion of whether alternative designs would be better. We expect cultural learners to first acquire the local practices, and occasionally experiment or modify them. At times this will mean that cultural learning will overrule their direct experience, evolved motivations, or reliably developing intuitions.

I think this passage highlights an important fact: culture is not merely a matter of *dissemination* of skills acquired individualistically, but a matter of the very constitution of the skills. (In one sense, this fact is trivial; but the truth is that people putting stress on individual cognitive capacities tend to underestimate it.)

However, as I have also already indicated, to account for the “gradual accumulation and recombination” we must accept that it is not just the mild versions of conflict that play an important role—culture and education is far from being only a *cooperative* enterprise. And here we come back to the concept of *rule*: the concept of a shared notion that something *ought-to-be*, that is transferred from individual to individual, and from generation to generation, by multifarious means of “enforcement”.

The conclusion reached here is, I think, fully in accordance with the conclusions of Castro et al. (2010, p. 351), who have proposed another epithet for our species—*Homo suadens*:

From this definition of imitation as a process of observational learning but not of replication of behaviors without evaluation, we contend that an increase in the efficiency of imitation is necessary but not enough to explain the transformation of primate social learning into a cumulative cultural system of inheritance in strict sense such as it happens during hominization. A key factor enabling such transformation was the fact that some hominids, which we call assessors, developed the capacity to categorize and to approve or disapprove their offspring learned behavior. This capacity to categorize learned behavior can be defined as the ability of an individual to categorize a given behavior by means of an

essentially dichotomous conceptual evaluative code: positive or negative, good or bad (...). This implies that an automatic mechanism for categorizing behavior used for individual learning is transformed into a conceptual categorization mechanism.

Consider a second-order pattern (such as forcing others to greet everybody in the morning). This is again a behavioral pattern, and it too could be forced by an appropriate enforcing behavior (a “third-order” pattern). If we could continue thus to infinity, we would have a behavior that is brought into being just by behavior, utterly without the support of genes; but of course this would be an infinite regress. Hence at *some* level the behavior must *not* be behaviorally enforced. It could be determined genetically, but if this was not present at the first level, then the question, as we already noted, would be why and how would evolution bring about a behavioral pattern in this detoured manner. Now the meta-pattern I am talking about above stops the regress at the second level: we do not need any third-order pattern to bring the second-order one into being, for it is brought into being in one sweep with the first-order pattern. Hence, being lauded if we greet others in the morning and shunned if we do not, we ourselves not only adopt the behavioral pattern of greeting others in the morning, but also adopt the tendency to laud those who greet others in the morning and shun those who do not.

Hence we are back at our original hypothesis. It is the peculiar ability to recognize a certain kind of coercion as not only something to which one is made to submit, but as an *ought-to-be* (as something to which everybody *should* submit and to which they *should* be made to submit, i.e. as something that is generally approvable), that gives us the raw material out of which our human culture and our human way of paragenetic spreading of behavioral patterns may arise. It is this which is so idiosyncratically needed to establish our *cultural niche*, the *milieu* of *Homo suadens ergo negotiatus ergo sapiens*.

7 Systems of Rules

It can be argued that rules are not quite specific to our species—that at least some other animals display collective behavioral patterns that appear to be “rule-governed”. I think this is largely a matter of terminology; but I would agree that a rudimentary version of this upgrade of us humans being shared by some of our animal cousins is a thesis open to discussion. But definitely a unique upgrade for us humans is our ability to develop *systems of rules*, which let us basically rebuild our natural niche.

Note the most basic problem with rules (and cooperation in general) from the viewpoint of evolution: a rule is not

useful unless it is generally accepted. (“I would give you my potato if I knew you would give me one of yours next time—but how do I know?”) This can be considered as a kind of holism; and its (not very surprising) solution seems to be a kind of bootstrapping—I start to follow the rule provisionally, and go on only if others reciprocate. Moreover, if I am a rudimentary rule-follower, each strap of the boot has to do not only with my personal *Bildung*, but rather also with forcing an order on the entire society.

For some rules, this kind of holism is the only problem—such a rule is clearly beneficial once in place, and the bootstrapping needed for its establishing is the only problem. However, besides the rules that require such substantial investment, there are some rules whose following comes almost for free. Take the rule of football forbidding any touching of the ball with one’s hands.³⁰ As there is little incentive to do so anyway, the costs of avoiding touching the ball are at most negligible (as in all cases of rules implementing pure conventions). Hence in this case there seems to be no grave problem in answering the question *how* rules could come into being. In contrast though, there may be a problem with answering the question *why* they came into being: there does not appear to be any obvious gain from not touching the ball, even if others reciprocate in the sense that they also avoid touching it.

One of the reasons for this situation is that these sort of rules involve, aside of the kind of holistic dimension mentioned above, also an *additional* kind of holistic dimension: they require, to be effective and functional, not only acceptance by many people, but also to be accepted together with many other rules. Football rules make no sense when taken one by one—they make sense only altogether. Hence in this case it is not individual rules that have a direct evolutionary *rationale* (like *tit-for-tat*); it is only their whole clusters.

But what may be the evolutionary rationale even for such clusters—what survival value might such a thing as football have? A possible response is to surmise that it is the rules immediately fostering cooperation that are truly key, and that other rules, such as those of football, are merely some kind of spin-off. From this perspective, it is “heavy-weight” rules, such as those of morality, that appear to be the *genuine* rules; the “light-weight” kind of rules being mere specters. However, I would like to point out that this perspective might obscure the important role of rules I urge here, namely their role in bearing the new, “software” way of spreading behavioral patterns. This is a role played by *all* rules, be they “light-weight” or “heavy-weight”.

From this perspective, the chasm between the rules of football and those of morality need not prevent us from

viewing all these varieties of rules as species of a single kind. Though clearly there are deep 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!”), the important features that both these enterprises share should not escape our attention. The common core we can recognize is the following. Rules regulate human conduct—they are applicable only to creatures that we hold to have a free will (i.e. that display behavioral flexibility exceeding a certain limit). Something is a rule only in so far as those governed by it are capable of potentially disobeying it. Rules make people behave in certain ways—enforce behavioral patterns. And further, and importantly, once the patterns became stable, they alter the environment in which we live and which channels our evolution.

Let us take note of the trivial fact that evolution is a matter of fitness *with respect to a specific environment*. Once people start to form communities, part of the relevant environment comes to be constituted by their peers. (This leads to the problem that fitness may start to be a matter of certain equilibria rather than simply of an optimization of features³¹). Moreover when rules start to play an important role within the communities (when what Knight, *ibid.*, calls the *rule of law* prevails), the tangible barriers of nature that channel evolution are increasingly replaced by artificial ones. We evolve due to 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. (This is, of course, not to say that we have somehow circumvented the bounds of nature—it is to say that these bounds affect us in a socially repacked and redistributed way.)

My suggestion now is that rules are crucial because of their ability to establish “virtual worlds”—virtual not in the sense of being unreal, but in the sense of owing their existence to the attitudes of people, namely to their *normative* attitudes that sustain rules capable of underpinning such virtual edifices.³² In this way, rules provide for a basic

³¹ See Maynard Smith (1998). It is also here where our behavior must become *strategic* in the sense of Sillari (2013).

³² I remember being struck, many years ago during the communist era in my country, by the realization that the whole oppressive machinery, whose presence seemed everywhere so tangible, was nothing but the product of attitudes of people—that the only thing making the almighty Regional Secretary of the Communist Party the unrestricted lord of human destinies was that this was what we all held him to be, and that if we all simultaneously stopped doing so, his majesty would collapse like a cardcastle. As Tomasello et al. (2012, p. 684) put it: “Social institutions are collaborative cultural practices with joint goals and standardized roles, with social norms governing how rewards are dispensed, how cheaters and free riders are treated, and so on. What is new about institutions is that they create new

³⁰ American readers should not be confused by *football* meaning what they call *soccer*.

alteration of the human niche (and hence effect what has been dubbed “niche construction”³³) and consequently of its evolution-fuelling features.³⁴ I suggest that thus rules provide for an acceleration of evolution, since they rob genetic replication of its exclusive right to promulgate patterns. Moreover, they provide for the evolution’s self-adjusting the barriers against which the selection that fuels it takes place.

Consider the development of computers. At first, the development (“evolution”) was a matter of the improvements of hardware. But once there appeared the idea of a multi-purpose hardware, 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 it is no longer guided directly by the tasks computers are to cope with (“environment”)—rather it is guided by the task to support, as best 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.

Similarly, we humans tend to ever more move into the “virtual” spaces from the “natural” one. The evolutionary pressures acting on us are ever more a matter of the “worlds” constituted by our rules than by the physical world governed by the laws of nature.³⁵ Sellars, as we saw, connects the emergence of rules with the emergence of behavior that is not “tied to the environment” in the way “most if not all animal behavior is”. Hence from this point of view, the difference is a matter of flexibility (in particular, with the emergence of behavior that may be “well characterized as free”). If we start by considering very simple organisms, the behavior of which consists of quite

rigid responses to their internal needs, then already the behavior of the higher animals like dogs or horses comes out as incomparably flexible—for *their* needs do not lead *directly* to behavior, but only produce states of the organism, “motivations” (emotions and the like), which move it to behavior not quite inevitably. As a result, various needs may compete for the resources of the organism producing a more or less optimized result.

We humans, from this viewpoint, have gained *one more* level of flexibility: namely “motivations” that are themselves, as it were, flexible. It seems that they are implemented, within us, merely “formally” waiting to be filled with a specific content. The “formal implementation” is achieved by means of a brand new kind of emotions, namely emotions that are usually called *social*.³⁶ Such emotions might be tied to specific targets, like the basic ones, in an inborn way, but mostly they are not. Such emotions as *pride* or *shame* get their targets only during phylogeny. And what is important, they may conspire to produce complicated social effects—rule following being the most important of them.³⁷ As the virtual, normative worlds in which we live can change quite rapidly (incomparably swifter than our natural world), rigid inborn “motivators” would be harmful—they could press an individual to behave in a way that would no longer be in accordance with its new environment. But such flexible ones that can be cast in the inborn form of social emotions by means of the materials of the external (mostly social) world would not have this shortcoming.

8 Conclusion

The behavioral patterns that constitute human societies are clearly less deterministically wired in our genes than is the case for any other species. Hence many of such patterns must be promulgated in some paragenetic way—via a “pedagogy” or a “culture”. Many authors studying this issue stress the concept of imitation, particularly the human ability to imitate others of our own species with an efficiency absent in other species. In this paper I have considered various proposals concerning what ingredients must be added to imitation to provide for the cultural spreading of patterns we witness in the case of our species. In accordance with the proposal of Castro et al. (2010) I argue that what is crucial is the disposition to approve and

Footnote 32 continued

status for individuals playing particular roles that everyone must respect,... These new statuses exist because and only because everyone agrees in common ground that they do; because institutions are especially clearly public, no one may ignore the new statuses by pleading ignorance of them (...).”

³³ See Odling-Smee (1996) or Day et al. (2003).

³⁴ See Peregrin (2011).

³⁵ It should also be noted that the claim goes beyond the frequent claim that culture constitutes a matrix that determines how we live, speak and think. The latter claim, in some circles, is so frequent that Tooby and Cosmides (1992) could classify it as “the Standard Social Science Model”. The view advocated here aims at revealing some of the mechanisms that provide for this *prima facie* odd picture, on which culture appears to act as a wholly mystical substance coming from nowhere and shaping the thinking of people—we try to indicate how the establishment of such a matrix can be seen as growing out of the emergent human capability of accepting rules—of instantiating the peculiar behavioral “meta-pattern” that leads to sustaining behavioral patterns across communities on the basis of beliefs, or of accepting *ought-to-be*’s as generating *ought-to-do*’s.

³⁶ See Barrett and Campos (1987); or Hareli and Parkinson (2008).

³⁷ As Frank (1988, p. 153) puts it: “People will try to avoid actions, motives, and qualities that make them feel afraid, sorry for those less privileged, anxious, bored, fatigued, or confused. The specific actions or circumstances that trigger those emotions will depend heavily on the cultural context. But the motivating emotions are always and everywhere the same.”

disapprove of others' behavior (and become sensitive to such approval/disapproval). This is primarily manifested in parents' teaching their offsprings, but then more crucially in the entire realm of social norms, where it provides for the complicated normative scaffolding of our societies.

I have pointed out that a radical step forward is taken when it becomes possible to merge the ability to force others to conform to some pattern with the ability to force them also to help force it. This merger, as I see it, comes with the emergence of rules and normativity, which then accompany our human species on its ascent to its current status. It has to do with people becoming sensitive to what others do and to the possibility of influencing them to eliminate "evil" doings, and then also becoming sensitive to others not being sensitive to the "evil". But what is important is that rules cannot be seen merely as an expedient in establishing (the rudiments of) morals. Rather, the evolution of rules provided for the toolkit usable for both morals and football, and totally changed the way we humans occupy our world.

This is to say that the emergence of rules may be the Rubicon, past which the natural environment, in which we humans lived like all other animals up to that point, started to be refined by, and amalgamated with, various rule-constituted superstructures. As a result, by now we live in a world that is much more "institutional" than "natural". We have lost the innocence of being straightforwardly tied to the environment. Moreover, the evolution of patterns hardwired in human brains have borne us an unexpected child: the spread of patterns carried by our rule-constituted culture.

References

- Axelrod R (1984) The evolution of cooperation. Basic Books, New York
- Barrett JC, Campos JJ (1987) Perspectives on emotional development II: a functionalist approach to emotion. In: Osofsky J (ed) Handbook of infant development, 2nd edn. Wiley, New York, pp 555–578
- Bickerton D (2005) Language first, then shared intentionality, then a beneficent spiral. *Behav Brain Sci* 28:691–692
- Binmore K (2005) Natural justice. Oxford University Press, Oxford
- Bowles S, Gintis S (2011) A cooperative species (human reciprocity and its evolution). Princeton University Press, Princeton
- Boyd R, Richerson PJ (2008) Gene-culture coevolution and the evolution of social institutions. In: Engel C, Singer W (eds) Better than conscious? Decision making, the human mind, and implications for institutions. MIT Press, Cambridge, pp 305–324
- Boyd R, Richerson PJ, Henrich J (2011) The cultural niche: why social learning is essential for human adaptation. *Proc Natl Acad Sci USA* 108:10918–10925
- Brandom R (1994) Making it explicit. Harvard University Press, Cambridge
- Castro L, Toro MA (2004) The evolution of culture: from primate social learning to human culture. *PNAS* 101:10235–10240
- Castro L, Castro-Nogueira L, Castro-Nogueira MA, Toro MA (2010) Cultural transmission and social control of human behavior. *Biol Philos* 25:347–360
- Cubitt RP, Sugden R (2003) Common knowledge, salience and convention: a reconstruction of David Lewis' Game Theory. *Econo Philos* 19:175–210
- Dawkins R (1976) The selfish gene. Oxford University Press, New York
- Day RL, Laland KN, Odling-Smee J (2003) Rethinking adaptation: the niche-construction perspective. *Perspect Biol Med* 46:80–95
- Donald M (1991) Origins of the modern mind: three stages in the evolution of culture and cognition. Harvard University Press, Cambridge
- Foucault, M. (1975): *Surveiller et punir: Naissance de la prison*, Gallimard, Paris; (English translation Discipline and Punish: the Birth of the Prison, Random House, New York, 1977)
- Frank RH (1988) Passions within reason: The strategic role of the emotions. Norton, New York
- Gabora L (2008) The cultural evolution of socially situated cognition. *Cogn Syst Res* 9:104–114
- Gendler TS (2008a) Alief and belief. *J Philos* 105:634–663
- Gendler TS (2008b) Alief in action (and reaction). *Mind Lang* 23:552–585
- Gergely G, Csibra G (2006) Sylvias recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In: Enfield NJ, Levenson SC (eds) Roots of human sociality: culture, cognition, and human interaction. Berg Publishers, Oxford, pp 229–255
- Guala, F. (2012) The normativity of Lewis conventions. *Synthese*. doi:10.1007/s11229-012-0131-x
- Hareli S, Parkinson B (2008) What is social about social emotions? *J Theory Soc Behav* 38:131–156
- Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: what is it, who has it, and how did it evolve? *Science* 298:1569–1579
- Henrich J, Gil-White FJ (2001) The evolution of prestige (Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission). *Evol Hum Behav* 22:165–196
- Joyce R (2006) The evolution of morality. The MIT Press, Cambridge
- Kelemen D (1999) The scope of teleological thinking in preschool children. *Cognition* 70:241–272
- Knight C (2008) Language co-evolved with the rule of law. *Mind Soc* 7:109–128
- Krebs JR, Dawkins R (1984) Animal signals: mind-reading and manipulation. In: Krebs JR, Davies NB (eds) Behavioural ecology: an evolutionary approach. Blackwell, Oxford, pp 380–402
- Kruger AC, Tomasello M (1996) Cultural learning and learning culture. In: Olson DR, Torrance N (eds) The handbook of education and human development: new models of learning, teaching and schooling. Blackwell, Oxford, pp 369–387
- Le S, Boyd R (2007) Evolutionary dynamics of the continuous iterated Prisoner's dilemma. *J Theor Biol* 245:258–267
- Lewis D (1969) Convention. Harvard University Press, Cambridge
- Maynard Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge
- Maynard Smith J (1998) Evolutionary genetics, 2nd edn. Oxford University Press, Oxford
- McElrath R, Boyd R, Richerson PM (2003) Shared norms and the evolution of ethnic markers. *Curr Anthropol* 44:122–130
- McKenzie AJ (2007) The structural evolution of morality. Cambridge University Press, Cambridge
- Meltzoff AN (1996) The human infant as imitative generalist: a 20-year progress report on infant imitation with implications for

- comparative psychology. In: Heyes CM, Galef BG (eds) *Social learning in animals: the roots of culture*. Academic Press, New York, pp 347–370
- Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314:1560–1563
- Odling-Smee FJ (1996) Niche construction, genetic evolution and cultural change. *Behav Process* 35:195–205
- Peregrin J (2010) The enigma of rules. *Int J Philos Stud* 18:377–394
- Peregrin J (2011) Creatures of norms as uncanny niche constructors. In: Hřfbek T, Hvorecký J (eds) *Knowledge value evolution*. College Publications, Cambridge, pp 189–198
- Poundstone W (1992) *Prisoner's dilemma*. Doubleday, New York
- Rakoczy H, Warneken F, Tomasello M (2008) The sources of normativity: young children's awareness of the normative structure of games. *Dev Psychol* 44:875–881
- Richerson PJ, Boyd R (2005) *Not by genes alone*. University of Chicago Press, Chicago
- Rochat P (2006) Humans evolved to become *Homo Negotiatus*... the rest followed. *Behav Brain Sci* 28:714–715
- Rouse J (2007) Social practices and normativity. *Philos Soc Sci* 37:46–56
- Sellars W (1949) Language, rules and behavior. In: Hook S (ed) *John Dewey: philosopher of science and freedom*. Dial Press, New York, pp 289–315
- Sellars W (1954) Some reflections on language games. *Philos Sci* 21:204–228
- Sillari G (2013) Rule-following as coordination: a game-theoretic approach. *Synthèse* 190:871–890
- Simon HA (1990) A mechanism for social selection and successful altruism. *Science* 250:1665–1668
- Skyrms Brian (1996) *Evolution of the social contract*. Cambridge University Press, Cambridge
- Skyrms B (2004) *The stag hunt and the evolution of social structure*. Cambridge University Press, Cambridge
- Skyrms B (2010) *Signals: evolution, learning, and information*. Oxford University Press, Oxford
- Spiekerman KP (2009) Sort out your neighbourhood: public good games on dynamic networks. *Synthèse* 168:273–294
- Tennie C, Call J, Tomasello M (2009) Ratcheting up the ratchet: on the evolution of cumulative culture. *Philos Trans R Soc B Biol Sci* 364:2405–2415
- Tomasello M (1999) *The cultural origins of human cognition*. Harvard University Press, Cambridge
- Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci* 28:675–735
- Tomasello M, Melis AP, Tennie C, Wyman E, Herrmann E (2012) Two key steps in the evolution of human cooperation. *Curr Anthropol* 53:673–692
- Tooby J, Cosmides L (1992) The psychological foundations of culture. In: Barkow J, Cosmides L, Tooby J (eds) *The adapted mind: evolutionary psychology and the generation of culture*. Oxford University Press, New York
- Tummolini L, Andrighetto G, Castelfranchi C, Conte R (2013) A convention or (tacit) agreement betwixt us: on reliance and its normative consequences. *Synthèse* 190:585–618
- Warneken F, Tomasello M (2009) Altruistic helping in human infants and young chimpanzees. *Science* 311:1301–1303
- West SA, Griffin AS, Gardner A (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J Evol Biol* 20:415–432
- Whiten A, Custance D (1996) Studies of imitation in chimpanzees and children. In: Heyes CM, Galef BG (eds) *Social learning in animals: the roots of culture*. Academic Press, New York, pp 347–370
- Wilson EO (1978) *On human nature*. Harvard University Press, Cambridge
- Wittgenstein, L. (1953): *Philosophische Untersuchungen*, Blackwell, Oxford; (English translation *Philosophical investigations*, Blackwell, Oxford)
- Woodcock S, Heath J (2002) The robustness of altruism as an evolutionary strategy. *Biol Philos* 17:567–590