Mindshaping and rules

Jaroslav Peregrin*

Department of Philosophy and Social Sciences, Faculty of Philosophy, University of Hradec Králové, Czechia

> *If we look at mankind inter-culturally, we find an amazing diversity. If we look at mankind within any one community/culture, we find an equally amazing discipline and restraint. Question: how can a species, genetically granted by Nature such remarkable freedom and licence, nevertheless observe such restraint, such narrowly defined limits, in its actual conduct? Man is born genetically free but is everywhere in cultural chains.*
>
> Ernest Gellner

Abstract. Our behavior is incomparably more flexible than that of any other animal species. The advantage is that we have become capable of surviving in different kinds of environments and withstanding even their abrupt changes; the disadvantage is that our behavior is *too* flexible, in that at every moment we can do not only useful and reasonable things, but also things that are useless and silly. Our evolution has solved this situation so that we complemented our "hardware" (genetic) freedom with a "software" (cultural) order. I suggest that the mechanism for this is based on rules; on the normative attitudes that we learn to assume towards their upkeep in interdependence with learning to respect them. A vital part of the process can be seen as mindshaping, as formatting one's mind to produce thoughts and actions that are easily interpretable, and also effective. It is not only that this enculturation passes hard-won capabilities from generation to generations; it is also that it formats the cognitions of the individual members of societies so as to chime with cognitions of other members.

* This work was supported by the Czech Science Foundation [G20-05180X].

Our behavioral plasticity

Let me, to begin with, consider an old and almost forgotten paper by Ernest Gellner (1989). "What the human species does share genetically," writes Gellner, "is an unbelievable degree of behavioural plasticity or volatility". This might seem unremarkable - even banal - but, in addition, Gellner is urging that this is something characteristic of our species. He remarks that the actual differences among people of different cultures cannot be a matter of genetics, but "what does seem genetically based in humankind is the plasticity, the volatility itself. All members of the species are endowed with it, and no other species possesses it." Gellner poses the following question: "How can a species, genetically granted by Nature such remarkable freedom and license, nevertheless observe such restraint, such narrowly defined limits, in its actual conduct?"

I think this kindles an intriguing perspective on humankind, and that Gellner is raising a profound question. Also, I think that currently we are in a position to provide, at least roughly, an answer. The fact that our behavior has such enormous flexibility in comparison with that of any of our animal cousins has clear advantages (many of which are probably why it has been selected for during evolution), such as our having become capable of surviving in different kinds of environments and withstanding even their abrupt variations. But there are also disadvantages: the vastness of our repertoire of behavior gives us, at any and every moment, the opportunity to attempt an awful lot of things, ranging from those that are useful and reasonable, to those that are useless and bizarre. This downside suggests that our behavior is *overly* flexible.

I will argue that evolution's response to this situation was to complement our "hardware" (genetic) freedom with a "software" (cultural) order. The "software" evolved precisely to counteract our behavior's excessive flexibility[1]. There are many tried and tested ways of coping with the world, and there are plenty of sensible ways of dealing with each other. But it would be disastrous if everybody were to seek these ways anew, by trial and error. If the environment changes, it may make sense to seek new ways; but unless this happens, it is more reasonable to stay with the tried ones.

How does culture manage to preserve and replicate useful forms of behavior? I think that the short answer is in terms of *rules*. Elsewhere I have discussed the nature of rules at length

[1] The comparison of our nature and culture with hardware and software of a computer may be deeper than meets the eye. I contend that human behavior's becoming so behaviorally versatile directly parallels computers' becoming universal. The thing is that computers became universal and extremely powerful because they could be turned, by means of software, into all kinds of special-purpose machines (like calculators, word processors, electronic diaries, game consoles etc.). However, without the software they would be inert. Cf. Rorty (2004).

(Peregrin, 2014; 2024a). An important point is that rules are not just explicit instructions - crucially, there exist rule-governed practices the rules of which are merely implicit to the practitioners' actions. However, this brings forth an important question: what is an implicit rule?

Implicit rules

When Wittgenstein (1953, §75) contemplated the nature of rules and of rule following, he came to face the fact that not every rule could be based on an interpretation. Imagine that we see a signpost, which tells us that by following certain red markings we will reach a town in an hour. To understand this, we must understand what is written on the signpost, and indeed the direction indicated by it. (Wittgenstein points out that even the latter is not self-explanatory – to understand it may require some enculturation.)

But if I must interpret the sign, it goes without saying that I must interpret it *correctly*. Hence, I must follow some rules of interpretation. If this rule following were to be again a matter of interpretation, we would need another rule and an infinite regress would be looming. To escape this, Wittgenstein thinks that we need some rules that are not explicit (in the form of signposts, linguistic instructions or whatever), the following of which does not involve interpretation: "Any interpretation still hangs in the air along with what it interprets, and cannot give it any support." (Wittgenstein, 1953, §198)

I propose a specific explanation of the nature of implicit rules (Peregrin, 2014; 2024a). Inspired by Brandom (1994)[2], I maintain that such rules exist in terms of *normative attitudes*, specific practical pro- and con- attitudes which are assumed by members of the society in question towards how they behave. It is important that the attitudes target only the *kinds* of behavior itself, and not its protagonists. Thus, if I oppose assault and support assistance, I must oppose/support them independently of how they concern me or any other specific person – whether it might be me who is assisted, or whether it is, say, Bob, who is engaged in assault[3].

If the attitudes to a kind of behavior align across a society, then we have a rudimentary implicit rule. Thus, e.g., if plus/minus all members of the society tend to oppose assault (not just assaulting specific persons or by specific brawlers, but assault as such), there is an implicit rule that assault is wrong, that one should not assault anybody. (Such a rudimentary rule will

[2] Due to the creative way I put Brandom's ideas to use, I do not want to make him responsible for anything I say, but I must acknowledge my deep indebtedness to him.

[3] But of course it may concern specific roles: a rule may be, e.g., that *nobody may assault anybody, with the exception of the chief, who may assault anybody* or that *nobody may assault anybody who does not deserve punishment* etc.

probably tend to evolve into something socially greatly more complex, perhaps with specific group of persons looking for violations of the rule and punishing it etc.)

A consequence of the fact that rules can exist via normative attitudes is that learning to follow a rule may be more similar to acquiring a skill, a know-how, than to acquiring explicit knowledge, a know-that. We learn to recognize and avoid the "social friction" that our actions sometimes elicit and that is a matter of others' negative normative attitudes; and we learn to act so that our actions elicit positive normative attitudes. Especially, then, we learn to live within our normative spaces by palpating the limits the rules pose for us and learning to deal with them as borders.

The role of normative attitudes

Imagine a driver weaving her way through a city duly observing all the rules of traffic. Imagine, then, a dog running through the same city and by chance doing this also in accordance with traffic rules. It would be hard to deny that there is a basic difference between these two cases. While the driver follows the rules of traffic in the sense of doing this intentionally, the dog only happens to comply with these rules without knowing it. The way we have just formulated the difference seems to indicate that it is located in the minds of the protagonists. While the driver has the rules, as it were, in her mind's eye, the same cannot be said of the dog. However, what exactly is it that is present in the mind of the human and absent from the mind of the dog?

It cannot be linguistic articulations of the rules: one can learn to follow the rules of traffic without ever verbalizing them, perhaps merely by trial and error (if one is lucky enough to survive the errors). It cannot be the knowledge of what to do in every possible situation: there are an infinite number of such situations. And in general, there doesn't seem to be a suitable mental object whose presence in the mind would render a behavior rule-governed and whose absence would signal the absence of the rule (Kripke, 1982).

The question at stake here is the following: how does a behavior governed by a rule differ from other kinds of behavior, especially from a behavior that is merely regular? And the answer I suggest is that we must seek the difference not somewhere "behind" or "beyond" the behavioral pattern we are investigating (*viz*. in the minds of its protagonists), but rather "beside" the pattern, in the broader context of the occurrence of this pattern. If an individual not only stops at red traffic lights, but also is angry with those who do not; if she not only observes speed limits, but also dutifully pays a fine if she happens to exceed them, and if she not only gives way, at crossroads, to those who have priority, but also counts on others doing the same, then she - unlike the dog - can be seen as following the rules of traffic.

It follows that without sufficient context there may sometimes be no answer to the question whether somebody is, or is not, following a rule. Does this mean that in such a case there is no difference at all? Not necessarily. There might be a difference which could perhaps be diagnosed by probing the person's brain: the driver might, for example, be thinking about the rules and control her driving accordingly (which might, in principle, be revealed by a brain scan). But it is also possible that she is driving absentmindedly, without even realizing that rules of traffic exist, and in such a case the answer might be unavailable.

Does this mean that rule following has nothing to do with minds? Not really. The point is that genuine rule following presupposes normative attitudes and assuming a normative attitude is not an occurrent event in the mind, but rather a relatively stable state. We do not stop to assume the attitude when we are not thinking about it, nor when we are asleep. Assuming such an attitude is something like supporting a sport team. If somebody goes to watch a certain team play, it is an indication of support, but we cannot be sure that he really supports it. We need a broader context.

Of course, it is possible to follow a rule without assuming the corresponding normative attitudes - one may be forced to follow it without accepting it. On the other hand, a rule, especially an implicit rule, cannot exist without the support of corresponding normative attitudes. If the rule is not to fade away, there must be a monitoring of its violations and a mechanism of rectification of possible deviations. This is a direction that the concept of normative attitudes leads us: a behavior is rule-governed only if it is generally measured by what *ought to be*, and it is so measured by the *behavior* of the protagonists of the allegedly rule-governed practice.

Social rules

There are two basic kinds of rules that are preserved (and perhaps upgraded) via the medium of our culture. The first kind are what von Wright (1963) calls *technical norms* or *directives*. They are *instrumental* rules, rules which help us achieve things of enduring value for us. Typically, they concern our dealings with nature: perhaps rules concerning how to build a kayak, how to quarry a deer, or how to construct a nuclear reactor. Many recent anthropologists have put stress on culture as the storage of such instructions (Boyd & Richerson, 2005; Henrich, 2015). But there is another important kind of rules, which are usually called *social*. These rules have to do with our cooperation and with our peaceful and productive coexistence with each other. These rules have to do not with effectiveness, but mostly with predictability, for predictability is the mother of peaceful coexistence. Mercier & Sperber (2017) argue that here is also the root of our all-important “game of giving and asking

for reasons" and consequently of our reasoning – that giving reasons germinated from situations when one does something unpredictable and wants to rectify this.

It may seem that just as the first kind of rules mentioned above help us operate successfully within the realm of nature, this other kind of rules help us operate successfully within the realm of our society. But this is not the whole truth. True, we *may* use social rules as directives to help us get what we want from our peers just like we use the instrumental rules to help us get what we want from the natural world. These rules *may* help us fruitfully navigate the landscape of the social world like the first help us navigate that of the natural one. However, what we call social rules also *create* the very landscape of the social world.

The natural world is not easy to deal with because it puts up resistance: many things we need or want are not easily obtained and we must invent clever ways to get them. The social world also wields resistance - but here the most intensive and the most systematic part of the resistance is normative - it is a matter of implicit rules carried by our attitudes, as well as of explicit rules, embodied in our codes and rule books.

Imagine a rule that that says that older people should be greeted. It can be used as a directive: you may, e.g., derive that if you want something from your peers, it is good to greet at least those of them who are older. But why does this work? Why is this a good course to navigate the social landscape? Because if you do not follow it, you may encounter resistance: you may be scorned by your peers and they may be less forthcoming to you. And why is that? Because they all accept the rule in the sense of evincing the corresponding normative attitudes.

Imagine, in contrast to this, the directive that you should not step on thin ice. Why is this a good course to navigate the natural landscape? Because if you do not follow it, you may encounter a resistance: the ice may break and you may drown. This danger is independent of what other people do; it is a matter of nature alone.

Hence, while a prototypical directive tells you what you should do (to get what you want) to overcome an objective resistance, in the case of social rules, the resistance that is to be overcome, is of your own making (of course, not of you alone, but of you and your peers collectively). When you learn to obey the rules, sooner or later you understand that you, as everyone else, are not only to submit to the rules, but also to come to take part in their enforcement, i.e. evince the normative attitudes that keep them in being[4].

Mindshaping

[4] Cf. Peregrin (2024b).

Zawidzki (2013) argues that an important part of human evolution was the development of a capacity which he calls mindshaping, i.e. "making human minds and behavior more homogeneous and hence easier to predict and interpret" (p. 29). According to him, this must have preceded the broadly discussed phenomenon of mindreading (Nichols & Stich, 2003; Lurz 2011), *viz*. the capacity of estimating the intentions (and the mental states more generally) of one's conspecifics.

Elsewhere (Peregrin, 2020) I pointed out that a crucial role within the process of mindshaping must have been played by our human invention of *rules* and here I am building on this finding. We have seen that humans have compensated for the excessive volatility of their behavior by establishing publicly sanctioned "channels" into which they steer each other's behavior. And it is the normative order of our social world which does this work.

Hence out of the vast array of things we could attempt, we generally restrict ourselves to those which are distinguished, or to doing things in distinguished ways. Importantly, it is not that we have to *decide* on the socially accepted courses of action, as the straight jacket of our society becomes our "second nature" to such an extent that we just act in the expected ways and deviate from them only for substantive reasons. This is the result of the fact that the normative attitudes, which permeate the entire society, act like walls: they reliably deflect us from certain courses of action letting us move only in certain directions.

Now what holds for our behavior also holds for our thinking: it too was conveyed into socially excavated channels. This may be hard to understand for those who feel that while mind causes behavior, behavior does not influence mind; but I think this conception is misguided. I side with Tomasello (2014, p. 38):

> The general process is thus that the young child imagines how some social interactant is comprehending or evaluating her, and then she uses this to socially self-regulate. Scaling up the sociality involved, children from about three years of age (but, needless to say, not apes) socially self-regulate on the basis of cultural structures—such as, prototypically, conventional and moral norms—that are based in cognitive processes of collective intentionality, what we may call normative self-governance. (...) Thus, from sometime during the late preschool period, young children self-regulate both their thinking and actions not just by how efficacious they will be in the current context (as do apes), and not just by how they will affect a particular person's thoughts or evaluations (as do younger children), but also by the perspective of how these will fit with the normative expectations of the social group. This process essentially constitutes the construction of a normative point of view as a self-regulating mechanism, arguably

the capstone of the ontogeny of uniquely human cognition (normative rationality) and sociality (normative morality).

Thus in putting up the normative infrastructure of a society we not only regulate behavior, but also thinking; or maybe we make people self-regulate.

Soft- instead of hard-wiring

Hence it would seem that rules allow us to retain a "soft" way of persisting in doing the "right" things. (Here "right" might mean tried and tested, or conforming to social conventions.) In simpler organisms, such ways may be "hardwired": the organism is simply set up so as to do things in these ways come what may. But we have seen that in our human case this would be a hindrance - the trajectory through evolution to which we have converged rests on flexibility. Therefore, we cannot make do with the "hardwiring"; but as abandoning it without a replacement would be equally disastrous as having it, we have developed the soft version[5].

Compare this with our emotions, the status of which is similar. Simple organisms may have hardwired connections between perception and behavior. In our case (and in the case of many other complex organisms) this would not work: we need a more flexible management of behavior. (Imagine that we are hungry and are attracted by a rosy apple on a nearby tree. If we were determined to assuage our hunger and to go for the apple come what may, we would not be able to avoid a predator which might turn out to be lurking in the tree. We need flexible enough management of behavior to allow us to flee, ignoring the lure of the apple.) This management is achieved by emotions (and later also by what are called executive functions, which can override emotions): they guide our behavior along the tried paths, but not rigidly. A more pressing emotion can trump a less pressing one, allowing the relatively optimal behavior to result from the interplay.

However, the more complex behavioral patterns cannot be produced by emotions alone. Here what is employed (over and above emotions) is a "collective memory" of the relevant community - the patterns are kept in place via rules, which are in turn upheld by normative

[5] Zawidzki (2018) also uses the metaphor of hardware/software. He points out that software, i.e. programs written in programming language, does not consist of descriptive statements, which aspire to represent the world (have the "language-to-world" direction of fit) but rather of commands, which attempt to make some changes in the world (have the "world-to-language" direction of fit), especially in the memory of the computer. Similarly, the "social software" is here not to equip individuals with tools to capture their inner world, but rather with tools that give the inner world its shape.

attitudes. This is a unique solution which has brought us, humans, to our enormously complex culture[6].

Nevertheless, normative attitudes, rules and culture cannot work independently of emotions; indeed, rules often build on the infrastructure of emotions. But the emotions are often insufficiently determinate to route them in any specific direction. Emotions like pride or shame must be filled with a specific content, which is what rules and culture are capable of doing. Hence such emotions can be seen as parts of a do-it-yourself kit which culture can use to build the normative network innervating a society.

It even seems that from this vantage point, mind may have developed precisely as a thing that can be soft-wired (or programmed, if you prefer), once the hard-wiring proved to be insufficient. Mind is the medium where you can imprint various things, including the optimal paths your behavior is to respect, without this needing to be permanent and indelible - mind can be updated.

But note also that mind alone is insufficient - crucial work must be done by a *community* of minds. We require the optimal paths to become imprinted into every new mind, and this necessitates a mechanism for the imprinting, viz. culture. Therefore, an individual mind must be equipped not merely by the imprint, but by the normative attitudes that effect it and that can be passed down the generations.

Zawidzki (2021) entertains similar ideas in the context of the relationship between individual metacognition and social cognition – his idea is that social cognition provides roles and categories which individuals employ for their metacognition, thus using the metacognition not for mapping an already existing cognition, but rather for giving it a socially constituted shape. Zawidzki writes (p. 6592):

> Person-level, linguistically expressible, metacognitive concepts are socio-cognitive tools that individuals acquire from their cultures, and that transform them as cognitive agents and cognitive objects, making them better at predicting others and easier to predict by others. Thus, they have both regulative and descriptive/predictive functions, but the latter depend on the former: we can use our metacognitive concepts to describe and predict each other's behavior because we have used them to regulate our own and each other's behavior in ways that make it easily describable and predictable in their terms.

Elsewhere (Peregrin, 2024a; forthcoming) I suggest that human life form is characterized by being structured into practices that are "specified by a system of rules which defines offices,

[6] See Peregrin (2024a).

roles, moves, penalties, defenses, and so on, and which gives the activity its structure" (Rawls, 1955, p. 33). I attempt to anatomize how such practices (and institutions) are constituted by rules and come to the metaphor of virtual arenas constituted by systems of rules, arenas within which we humans have now come to dwell and which permit us to engage in actions that would be unavailable to us should we stay outside of them.

Being born into a human society, we soon enter, during an early stage of our upbringing, the most important of these arenas (such as the arena of morality and that of language) and learn to live within them. Also, we learn how to visit many other arenas (e.g. those of various games); and later we learn how to help maintain the arenas and possibly how to help create new ones. The arenas make it possible for us to live our "unnatural" lives (as, instead of living within the milieu of nature alone, we largely live them within the "virtual" milieus of our own making). And this whole labyrinth of normative spaces is rendered possible by our submitting ourselves to the discipline of rules and having our minds (benignly) shaped.

Conclusion

Genetic hardwiring of behavioral patterns is a mighty tool; however, once the complexity of the organism in question increases beyond a certain degree, it can become a hindrance. A more flexible solution is needed; and in our human case, it is the soft-wiring of behavioral patterns by means of structures preserved via social cognition. I suggest that the mechanism in which such enculturation works is based on rules; on the normative attitudes that we learn to assume in interdependence with learning to respect them. I also suggest that a vital part of the process can be seen as mindshaping, as learning to steer one's thoughts and actions into socially sanctioned channels.

We humans are social animals; and we have driven our sociality to an unprecedented form of cooperation and mutual dependence. This, on the one hand, is treated as a matter of course, while, on the other hand, it is often not thought through to its consequences. In particular, there is a unique interplay between social cognition and individual cognitions of members of the society. It is not only that the society stores hard-won capabilities, which it instills in new members; it is also that it formats the cognitions of the individual members so as to chime with cognitions of other members.

Gellner's paper that I quoted at the beginning of this article closes as follows (p. 525):

> My argument has been that genetic under-programming must have been linked to the presence of a compensating system of cultural/linguistic restriction. These cultural systems, and systems of coercion, have complemented each other in

> diverse ways at different stages. The volatility must obviously have had its own genetic preconditions, so that our volatility, and our endowment with compensating talents and propensities, must have arrived jointly. The consequence has been the emergence of a species in whose life both social or semantic transmission and institutionalized coercion have become far more important than genetic mutation, making it possible for change to be astonishingly rapid.

To these words of my virtual compatriot[7], I wholeheartedly subscribe.

References

Boyd, R. & Richerson, P. J. (2005): *Not by genes alone*, Chicago: University of Chicago Press.

Brandom, R. (1994): *Making it explicit: Reasoning, representing, and discursive commitment*, Cambridge (Mass.): Harvard University Press.

1. Gellner, E. (1989): 'Culture, constraint and community: semantic and coercive compensations for the genetic under-determination of Homo sapiens sapiens'. *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans* (ed. by P. Mellars & C. B. Stringer). Edinburgh: Edinburgh University Press.

Henrich, J. (2015): *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*, Princeton: Princeton University Press.

Kripke, S. A. (1982): *Wittgenstein on rules and private language: An elementary exposition*, Cambridge (Mass.): Harvard University Press.

Lurz, R. W., ed. (2011): *Mindreading Animals: The Debate over What Animals Know about other Minds*, Cambridge (Mass.): The MIT Press.

Mercier, H. & Sperber, D. (2017): *The enigma of reason*, Cambridge (Mass.): Harvard University Press.

Nichols, S. & Stich, S. P. (2003): *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford: Oxford University Press.

Peregrin, J. (2014): *Inferentialism: why rules matter*, Basingstoke: Palgrave.

[7] Like myself, Gellner spent his early years in former Czechoslovakia (though I lived there significantly later than him). The onset of Nazism in Germany made his family leave for Britain.

Peregrin, J. (2020): 'Normative Mindshaping and the Normative Niche'. *Groups, Norms and Practices* (ed. L. Koreň et al.), pp. 85–98, Cham: Springer.

Peregrin, J. (2024a): *Normative Species: How Naturalized Inferentialism Explains Us*, New York: Routledge.

Peregrin, J. (2024b): 'Normativity between philosophy and science'. *Philosophical Psychology* 37(5), 1215-1239.

Peregrin, J. (forthcoming): 'Inside Human Practices', *Practices in Legal and Social Theories* (ed, by M. Dybowski et al.), New York: Routledge.

Rawls, J. (1955): 'Two Concepts of Rules'. *Philosophical Review* 64(1), 3–32.

Rorty, R. (2004): 'The brain as hardware, culture as software: Symposium: Vincent Descombes, The Mind's Provisions'. *Inquiry* 47(3), 219–235.

Von Wright, G. H. (1963): *Norm and Action*, New York: Humanities Press.

Wittgenstein, L. (1953): *Philosophische Untersuchungen*, Oxford: Blackwell; English translation *Philosophical Investigations*, Oxford: Blackwell, 1953.

Zawidzki, T. W. (2013): *Mindshaping: A new framework for understanding human social cognition*, Cambridge (Mass.): MIT Press.

Zawidzki, T. W. (2018): 'Self-Interpretation as Software: Toward a New Understanding of Why False Self-Conceptions Persist'. *Third-Person Self-Knowledge, Self-Interpretation, and Narrative* (ed. P. Pedrini and J. Kirsch), pp. 115–144, Cham: Springer.

Zawidzki, T. W. (2021): 'A new perspective on the relationship between metacognition and social cognition: Metacognitive concepts as socio-cognitive tools'. *Synthese* 198(7), 6573–6596.