Australasian Journal of Philosophy

ISSN: 0004-8402 (Print) 1471-6828 (Online) Journal homepage: https://www.tandfonline.com/loi/rajp20

Kim Sterelny, Thought in a Hostile World: The Evolution of Human Cognition, Oxford: Blackwell, 2003, pp. xi + 262, £50 (cloth), £16.95 (paper). Friendly Thoughts on the Evolution of Cognition

David Papineau

To cite this article: David Papineau (2004) Kim Sterelny, Thought in a Hostile World: The Evolution of Human Cognition, Oxford: Blackwell, 2003, pp. xi + 262, £50 (cloth), £16.95 (paper). Friendly Thoughts on the Evolution of Cognition, Australasian Journal of Philosophy, 82:3, 491-502, DOI: 10.1080/715691176

To link to this article: https://doi.org/10.1080/715691176

Published online: 08 Apr 2010.

Submit your article to this journal

Article views: 207

View related articles

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rajp20

Australasian Journal of Philosophy
Vol. 82, No. 3, pp. 491–502; September 2004

CRITICAL DISCUSSION

Kim Sterelny, *Thought in a Hostile World: The Evolution of Human Cognition*, Oxford: Blackwell, 2003, pp. xi + 262, £50 (cloth), £16.95 (paper).

FRIENDLY THOUGHTS ON THE EVOLUTION OF COGNITION

David Papineau

Thought in a Hostile World covers an admirable range of topics. It falls into three parts. In the first Kim Sterelny discusses the general evolution of animal cognition. The second considers what specific forces may have guided the further evolution of the human intellect. The final part then reflects on alternatives to the currently fashionable view that the human mind is massively modular. Along the way Sterelny has many interesting things to say, on subjects as varied as spatial navigation, imitation, rat intelligence, group selection, niche construction, neural plasticity, folk biology, and the frame problem.

Not surprisingly, given this varied subject matter, there is no sound-bite sized thesis that summarizes the book. There is too much going on for any quick précis. Yet this is the first book I would recommend to any graduate student who wants to learn about comparative and evolutionary approaches to cognition. Sterelny provides an excellent introduction to nearly every topic of current philosophical interest in this area. In many cases this requires familiarity with a complex body of empirical material, but throughout Sterelny reviews the issues effortlessly and with insight.

Still, even if there is no quick digest of Sterelny's position, he does have a distinctive point of view. Unlike the MIT-led orthodoxy of the last few decades, he takes evolution seriously, both as a possible source of information about the modern mind, and as a topic of interest in its own right. At the same time, he differs markedly from the self-styled 'Evolutionary Psychologists', in his readiness to consider alternatives to modular and nativist hypotheses. To my mind, these are both eminently sound attitudes, even though they are not widely shared. I find it hard to think of another contemporary writer who displays such good judgement across such a range of topics.

While I am in emphatic agreement with Sterelny's overall approach, there are plenty of more specific points on which I am not entirely convinced by

his position. There is no question of commenting on all of these in the space available. So in what follows I shall concentrate almost entirely on topics from his first section, on the general evolution of animal cognition. Still, before turning to this, I would like briefly to enter one caveat about an important general issue from later in the book.

This concerns Sterelny's discussion of nativism and modularity in Part III of the book. As I said, Sterelny is resistant to 'Evolutionary Psychology', and its central contention that the human mind consists of nothing but innate and encapsulated cognitive 'modules'. He is prepared to concede that language probably rests on such a module. But he thinks that this may be a special case. When it comes to folk psychology, folk biology, and so on, Sterelny sees no reason to posit similar genetically-driven cognitive mechanisms. If anything makes humans cognitively distinctive, says Sterelny, it is that they have been selected for developmental plasticity, not rigidity. This is not to deny that most distinctively human cognitive powers are inherited from prior generations. But, language apart, this inheritance is non-genetic. Humans cumulatively construct complex cultures, replete with tools, stories, behavioural exemplars, as well as explicit theories. As a result, their children grow up in uniquely information-rich cognitive environments. These environments, rather than genetic selection, are enough to explain human cognitive sophistications.

If I had to choose, I would agree that this picture contains more truth than hard-line Evolutionary Psychology. However, I was uneasy about the extent to which Sterelny presented the issue as an either/or matter: he seems assume that human cognitive traits are *either* innate *or* cultural. This is not the place to start analysing the notion of innateness, but one interesting question we can always ask is whether a trait depends in significant part on genes that have been selected specifically because they facilitate that trait. I didn't find anything in Sterelny's arguments to make me doubt that human cognitive abilities like folk biology and folk psychology will in this sense have substantial genetic underpinning.

Baldwin-like considerations are obviously relevant here. Let us assume that human children grow up surrounded by rich folk biological and psychological cultures. To the extent that their acquisition of these cultures is important, shouldn't we expect strong selection pressures for genes that make human children learn these cultures quickly and reliably? Sterelny mentions the Baldwin effect in connection with language, but queries whether it is significant for these other cognitive abilities, on the grounds that humans continue to develop these other abilities into adolescence. But, again, this is not an either/or matter. It could be that there has been significant selection of genes to foster folk biology and psychology, but that these genes provide only limited support, helping to ensure that certain basic

categorizations are picked up quickly, but leaving it solely to longer-term learning and (perhaps locally variable) culture to fill in the details. This would make folk biology and psychology *less* innate than language, but wouldn't mean they had no genes to call their own. (By the same coin, even if Baldwinization has meant that language is by comparison *more* innate, this is no reason to deny that cultural factors continue to play a deep role in its development.)

When the issue is posed in these terms, I suspect that Sterelny would not wish to disagree. He allows, for example, that some innate perceptual sensitivities, for colour vision, say, may have been selected specifically because they help us cotton onto folk biological categories. Moreover, he views human cultural environments as a special case of 'niche construction', and endorses the idea that niche construction changes the genetic selective landscape. Given this, perhaps his general rejection of 'nativist' views is merely intended to distance himself from those Evolutionary Psychologists who take nativism to imply that cognitive traits will grow willy-nilly, like fingers, provided basic requirements of nutrition are met. Still, it seems to me that viewing the topic in this way already concedes too much to the Evolutionary Psychologists. For it comes close to accepting that, if a trait is interestingly 'innate' at all, then it must be innate in the way that the Evolutionary Psychologists have in mind. We should leave room for the possibility that significant suites of genes have been selected specifically to foster traits like folk biology and psychology, while at the same time allowing that the ontogeny of these traits also depends crucially on non-genetically inherited and possibly locally variable cultural environments.

Let me now turn to some of the general issues about the evolution of animal cognition raised in Part I. In general, I found this section highly stimulating. Sterelny asks questions about the basic mechanisms of cognition that are ignored by nearly all other writers in this area. However, here too I felt that some topics were being treated in an overly dichotomised way. Where Sterelny sees clear rifts in the evolution of cognition, it seemed to me there may be more continuity. I shall consider three issues from Part I: decoupled representations, preferences, and causal knowledge.

I. Decoupled Representations

A central theme in Part I is the evolution of 'decoupled representations'. Decoupled representations contrast with what Ruth Millikan calls 'pushmi-pullyu' states—that is, states which are triggered by environmental cues but which are capable of prompting only one kind of behaviour. The frog's fly-catching response would be a case in point—the sensory states that trigger

head-turning and tongue-snapping never lead to any other behaviour but this. By contrast, as Sterelny explains, decoupled representations, are 'belief-like': they have no behaviour to call their own, but rather help to guide different behaviours as appropriate in different contexts.

I agree with Sterelny that the evolution of decoupled representations marks a crucial stage in the evolution of cognition. But I have doubts about the way he fills out the notion. For Sterelny, the crucial criterion is simply 'response breadth'. A given representation is decoupled to the extent in prompts different behaviours in different cases. To take Sterelny's own initial example of 'response breadth', a vervet monkey's recognition of a martial eagle can variously lead to concealment, to readiness to fight, or merely to higher alertness, depending on other circumstances.

However, as Sterelny notes in a footnote (crediting Matteo Mameli and Peter Godfrey-Smith), 'response breadth' in this sense can be read as merely indicating that the monkey is making finer perceptual discriminations than simply 'martial eagle'. Perhaps the monkey is variously registering 'martial-eagle-in-middle-distance', 'martial-eagle-imminent', 'martial-eagle-far-away', and then performing the single response that is tied to these finer representations. Looking at things this way, it is unclear that the finer representations have become decoupled from specific responses after all.

Sterelny's answer is that at least the representation of 'martial eagle' per se has becomes decoupled from any specific response. This is fair enough, but I wondered why Sterelny did not adopt a rather different strategy for identifying decoupled representations. Why not say that a representation is decoupled when it leads to different responses in combination with different *motivational states*? This promises to pick out a category of 'belief-like' representations whose 'decoupling' is absolute, and does not disappear when they are conjoined with further belief-like states.

Sterelny has things to say about the evolution of motivational states themselves—and in particular about the contrast between drives and preference—and I shall comment on this in a moment. But we don't need any detailed analysis of motivation at this stage. Define a motivational state as one whose function is to produce some definite effect (food, warmth, alcohol, music, whatever). Then we can identify belief-like states as those which have no effects to call their own, but which rather are slave to different motivational states. Their function is not to yield any particular outcome, but rather to inform the motivational state, so to speak, about current circumstances, and hence about which response will secure its end in those circumstances. To take an example, visual recognition of a fruit tree will lead an animal to shelter in its shade when it wants to cool down, but to climb the tree in search of food when it is hungry. From this perspective,

then, 'belief-like' states are distinguished by their general-purposiveness—that is, by their availability to foster the pursuit of many different biological ends.

Sterelny suggests that decoupled representations may have emerged relatively late in evolution. He considers spatial and causal knowledge as possible species of decoupled representation that may be widespread throughout the animal kingdom, but points out that there is no clear evidence that these representations support a broad range of different behavioural responses, and conjectures (notwithstanding his initial example of vervet monkeys and martial eagles) that genuine decoupled representations may not have emerged until the social tracking of cognitive states in great apes. From the point of view of the suggestion I have just made, however, we might expect decoupled representations to be much more widespread. They will be present as soon as animals can use a given representation (the visual representation of a tree, say) to guide different behaviours when in different motivational states.

Still, it should not be taken for granted that even decoupling in this sense is widespread. For many animals, any given perceptual pathway may be tied to some specific behaviour, and not be available to serve different motivations. Even an animal who can use a tree both for shade and fruit may be responding to different visual aspects of the tree (perceived shadow leads to shade-seeking, perceived fruit leads to climbing). Frogs seem to work rather like this. The visual information that guides the frog's prey-catching behaviour is quite different from the visual information that guides its obstacle-avoiding behaviour, and different again from the visual information that leads it to jump away from looming threats. Lesions of the frog's optical system can dissociate these different abilities [Milner and Goodale, 1995, sect. 1.2.2.]. To the extent that each visual representation is dedicated to some specific behavioural response, the frog's representations are not decoupled.

What about animals that can learn? (Sterelny takes it, throughout the book, that a wide range of animals are capable of basic kinds of conditioning.) Won't we have decoupled representation in any animal that can be instrumentally conditioned to treat some given perceptual representation as a stimulus for different behaviours in different motivational circumstances? Maybe so, but we need to go slowly here. There are different types of possible instrumental conditioning. Simple stimulus-response (S-R) conditioning leads an animal to associate a stimulus with a behavioural response, but without registering, so to speak, what outcome the response is good for. In response-outcome (R-O) conditioning, by contrast, the animal also registers which outcome is produced by the relevant response. The most salient

operational difference is that, with R-O conditioning, the animal will not perform the response if it is not motivated to seek the outcome, but not so with S-R conditioning. For example, consider an animal whose lever-pressing when the bell rings has been rewarded with hunger-satisfying food: R-O conditioning will mean that it will only subsequently lever-press after the bell when it is hungry, but S-R conditioning will lead it to lever-press whenever it hears the bell, whether it is hungry or not.

The upshot of this distinction is that only R-O conditioning implies decoupled representations in the sense I have suggested. This is because only R-O conditioning yields gives rise to behaviour that is sensitive to motivational state, and thus to animals who will perform *different* behaviours in response to a given perceptual stimulus in *different* motivational states. We can imagine an R-O conditioned rat, say, who presses the lever when the bell rings if it is hungry, but who pulls a chain when the bell rings if it is thirsty. Here the rat's aural representation of the bell-ring will be decoupled in my suggested sense: it prompts different behaviours given different motivational states.[1]

How widespread is R-O conditioning? To the extent that this relatively sophisticated form of learning is rare in the animal kingdom, it will support Sterelny's underlying thought that decoupled representations only emerged comparatively late in the evolutionary story. However, it is not clear that Sterelny himself can appeal to this line of argument. This is because he implicitly commits himself to the assumption that R-O learning is present in a wide range of mammals—or at least, to be more specific, that it is present in rats. In discussing an important range of experiments on rat learning by Tony Dickinson, Sterelny resists Dickinson's high-level cognitive explanation of his rat data, precisely by insisting that an alternative associationist explanation will still be possible, provided we allow the rats a relatively fine-grained range of drives, and suppose that their learned responses are sensitive to the presence or absence of these drives [85-6]. This seems to me quite the right thing to say about Dickinson's rat data: they can indeed be explained in terms of R-O learning, without positing high-level cognitive 'reasoning'. But, insofar as this is right, then rats will also be capable of significant representational decoupling, for the reasons I have explained.

[1] S-R conditioning, by contrast, cannot yield this kind of decoupling, since whatever behaviours it gives rise to will be insensitive to motivational state. For an animal subject only to S-R conditioning, a given stimulus will automatically prompt a given response, and so won't be dissociable from the response by variations in motivational state. (Interestingly, it is arguable that even S-R conditioning will standardly give rise to decoupling in Sterelny's preferred sense of 'response-breadth'. Even an S-R animal can learn to do one thing when it sees a far-away fruit tree, and to do something else when it sees a fruit tree nearby.)

II. The Descent of Preference

Sterelny's discussion of Dickinson's rats occurs in the course of his chapter on the evolution of 'preferences'. As Sterelny sees it, preferences are distinct from drives. Preferences involve representations of aimed-for changes in the external environmental, where drives simply contribute to internal systems of homeostatic regulation. Moreover, as with decoupled representations, Sterelny takes preferences to arrive relatively late on the evolutionary scene, perhaps only in the hominid lineage.

I thought that this chapter raised many important issues. But I must say that I found the basic distinction between preferences and drives less clear-cut than Sterelny does. Perhaps the best way to focus this issue is to consider the four ways in which Sterelny takes preferences to mark an evolutionary advance on drives. He says that preferences (1) liberate motivation from immediate affective reward, (2) can tie motivation to the external environment, (3) allow animals to learn new wants, and (4) allow behaviour to be responsive to a ranking of motivations.

I didn't see why any of these criteria draw the kind of distinction between drives and preferences that Sterelny has in mind. Let me take them in turn.

A. Affective Reward

Sterelny's thought is that the satisfaction of a basic drive like thirst, say, involves some kind of immediately rewarding sensation, of a kind which is not associated with a higher desire, say, to read Middlemarch again. However, it is not clear that higher desires can really be dissociated from 'affective reward' and still retain any motivational 'push'. This seems to be one lesson of Antonio Damasio's work on emotions. In *Descartes' Error* (1994), Damasio describes patients (the most prominent is 'Elliott') who have lost abilities to react 'emotionally' (cf. 'affectively') to real or imagined scenarios. To the extent that these reactions are lost, so is the patient's ability to decide between alternative courses of action. Damasio explains how 'Elliott' would reflect on everyday choices for up to half an hour without being able to decide what to do. It seems that even the kinds of higher preferences that Sterelny is thinking of need to retain links to affective responses, if they are to have any motivational force. If reading Middlemarch, or envisaging reading it, does not make you feel good, then you won't be moved to do it.

B. The External Environment

Here the idea is that preferences aim for changes in the external environment, like a new roof on your house, say, while drives simply register some internal

metabolic index, like blood sugar level. However, it seems to me that even some basic drives are targeted on external circumstances, whatever internal mechanism regulates them. Take the male sex drive. Surely the target of this drive is to impregnate the female (or her eggs, if fertilization is external), not just to lower the level of the hormones involved in sexual arousal. After all, the whole point of male sexual activity is to effect this change in the outside world. I take this example to show that, even with the most basic drives, it is in principle possible to distinguish their targets (their satisfaction conditions) from the internal mechanisms that turn them on and off. In some cases, true, the satisfaction condition may coincide with the equilibrium state of the internal mechanism—for example, temperature regulation systems presumably both aim at a target temperature and are activated by deviations from it. But in other cases, like the male sex drive, the targeted satisfaction condition may be quite distinct from the internal regulatory system.

C. Learning New Wants

I agree with Sterelny that the ability to acquire new wants is a crucial stage in the evolution of cognition. But once more I doubt that this marks any clear division between drive-based systems and more advanced cognition. The phenomenon known as *secondary reinforcement* is relevant here. Standard associative learning theory tells us that some circumstance that is not initially rewarding to an animal can come to acquire a positive motivational value if it is experienced as preceding something that is already rewarding: the animal comes to value circumstances that it experiences as precursors or means to things it already values. For example, suppose that an animal habitually passes some landmark on its way to feeding. Then it will come to value passing the landmark in itself, as will be shown by this new reward's ability to reinforce other behaviours, even when it is not followed by feeding. (Of course, continued experience of the landmark not being followed by food will reverse the process, and render the landmark neutral in affect once more [Gross 1996: 164].) Now, I have no doubt that this kind of motivational learning is of great evolutionary importance. After all, it means that animals can come to value anything that they can recognize, and are not limited to some basic stock of innate drives. Still, it doesn't look like something that marks out some special category of primate thinking, given that is widespread among the many birds and mammals that display secondary conditioning.

D. Ranking of Wants

Sterelny is particularly interested in behaviour which simultaneously displays the influence of different motivations. Mating chimpanzees will

sometimes suppress copulation cries, thus showing they are motivated to avoid detection as well as to have sex. Sparrows will recruit others to join in feeding at a risky-looking site, so manifesting a concern to avoid predation alongside their hunger. Sterelny observes that such modulation of behaviour would not occur in a 'winner-take-all' motivational structure in which the strongest drive simply eclipses any other active drives. This is an interesting point, but whether it marks out an evolutionarily distinctive kind of motivational system depends on whether simpler animals do in fact have 'winner-take-all' motivational structures and so fail to display the kind of 'portmanteau' behaviour that Sterelny is interested in. It is not clear to me why *any* animals should be limited in this way. After all, there seems no need to hand over all behavioural control to one among various active motives unless these motives are pressing for *incompatible* behaviours—and in such cases there seems no option to the 'winner-take-all' strategy. Conversely, if you *can* eat and watch out for predators at the same time, then there will be no need for your motivational system to decide whether your hunger or your anxiety should direct behaviour—they both can. Given this, I would expect all animals to be capable of displaying behaviour that shows the simultaneous influence of multiple motivations, simply because such 'portmanteau' behaviour would seem to be the default option whenever it is possible. So I have doubts about Sterelny's suggestion that such behavioural modulation marks a significant step in the evolution of motivation.

Let me sum up this discussion of preferences. I agree that there is no doubt that apes, and in particular human beings, are capable of a wider and more interesting range of preferences than seagulls, say. But I think that this is a matter of degree rather than of kind. None of Sterelny's four putative distinctive features of preferences seems to me to mark a definite line between preferences and drives. True, the third feature does at least serve to distinguish learned motivations from innate ones. But, as I said, the learned acquisition of motives will be common to any creatures capable of secondary conditioning, which I take to include rather more than Sterelny's category of creatures with preferences.

Of course, as I observed, any animals that are capable of recognizing a wide range of different circumstances will be able, via secondary conditioning, to acquire motivations targeted on any of those circumstances. And this does means that humans and their close relatives are capable of valuing a far wider range of things than other animals. However, given the points I have made, it would seem wrong to explain this in terms of humans and their close relatives having a special kind of motivational structure. Rather the difference simply lies in their greater representational powers. All the motivational structures we have considered, and the power to extend them by learning, are

common to many animals. The reason the human lineage can learn to value more is simply that it is capable of representing more.

III. Causal Knowledge

If I were looking for significant features of 'higher' cognition, I would concentrate on an issue that Sterelny touches on only in passing. In the course of discussing 'decoupled representations', Sterelny considers the sensitivity of rats to causal relationships [37–8]. Experimental work, by Tony Dickinson again, has shown that rats are sensitive to the difference between genuine causal relationships and spurious patterns. Suppose rats receive food after pressing a lever. They will only continue to press the lever when they have reason to think that the food really does *depend* on their lever pressing. Thus they won't continue to lever-press if the food also arrives as frequently when they *don't* press the lever. However, this effect is reversed in regimes where those other food deliveries are preceded by some yet further event, like a brief flash. In this latter case, the rats *will* learn to lever-press, apparently judging that the food can *either* be caused by their lever pressing *or* by the brief flash.

Sterelny argues that even this sophisticated sensitivity to causal facts doesn't yet amount to 'decoupled representation', on the grounds that the rats have only been shown to do one thing with their causal information—namely, press levers. Dickinson's experiments give us no reason to suppose that the rats can use this causal information to inform other kinds of actions. I agree entirely about this. The rats' causal knowledge strikes me as a paradigm of 'undecoupled' information.

However, if seems to me that Sterelny would have done well to have considered a deeper explanation for the rats' limitations. Compare the rats' causal knowledge with other kinds of representation discussed above—representations of particular current facts like martial eagles, fruit trees, bell rings, and so on. In all these latter cases, the representations will presumably be embodied in a passing modification of some perceptual system. By contrast, the rat's causal knowledge is likely to be embodied *procedurally* rather than perceptually—the information will simply reside in the rat's behavioural disposition to lever-press in comparable situations when hungry. If this is right, it is inevitable that this information cannot be decoupled from the lever-pressing behaviour. To the extent that the rat's information (that lever-pressing yields food) *consists in* its inclination to lever-press when hungry, there is no question of this information guiding any other behaviour than lever-pressing.

If you ask me, I would say that most animals store most of their general causal information in just such behavioural dispositions. Where knowledge

of particular facts about the current (and recent) environment will be registered perceptually, causal knowledge will standardly be embodied in dispositions to perform given behaviours in given circumstances in pursuit of given goals. If this is right, then it will impose severe architectural limitations on the ways in which animals can use their causal information. True, there are many interesting kinds of flexibility that even procedurally embodied causal information will allow. For example, classical Pavlovian associations between stimuli will allow animals to transfer embodied causal information from one circumstance to another, and secondary reinforcement will allow animals to 'chain together' different items of embodied causal information [Papineau 2001; 2003]. But a strong constraint will remain: to the extent that causal information consists in dispositions to specific behaviours, animals will only be able to manifest this information in just those behaviours.

I am not sure how far *all* non-human cognition is limited in this architectural way. (If it were, it would do something to explain the striking inability of non-human primates to understand simple causal mechanisms [46-7].) However, the empirical extent of this kind of architectural limitation is not my main point here. Rather, I simply want to observe that, if we are interested in the evolutionary emergence of decoupled representations, one important issue to focus on is the shift from the procedural embodiment of causal information to the explicit representations of such information and its consequent availability to open-ended inferential manipulation. Maybe this shift has occurred in a number of animals, or maybe it is peculiar to our recent hominid lineage. But, in any case, we need to understand how it occurred, and how the new inferential abilities related to older modes of behavioural control.

In final conclusion, let me stress once more that my differences from Sterelny reflect a wide measure of underlying agreement. I have found it a pleasure to engage with him on points of detail precisely because I am so persuaded of his overall approach. I may not agree with all his answers, but he is certainly asking all the right questions. This book should be required reading for all students of the evolution of cognition.

King's College London

REFERENCES

Damasio, A. 1994. *Descartes' Error*, New York: Putnam.
Gross, R. 1996. *Psychology*, 3rd edn, London: Hodder and Stoughton.

Milner, A. and M. Goodale 1995. *The Visual Brain in Action*, Oxford: Oxford University Press.

Papineau, D. 2001. The Evolution of Means-End Reasoning, in *Naturalism, Evolution and Mind*, ed. D. M. Walsh, New York: Cambridge: Cambridge University Press: 145–78.

Papineau, D. 2003. Human Minds, in *Minds and Persons*, ed. A. O'Hear, Cambridge: Cambridge University Press: 159–83.