

The Biological Basis of Cultural Transmission  
Review of *Thought in a Hostile World* by Kim Sterelny  
Oxford: Blackwell, 2003

*British Journal for the Philosophy of Science*, 57 (2006), 259-266  
doi:10.1093/bjps/axi162

Nicholas Shea

Once someone hits upon a good idea, others can learn it from them with ease, and develop it further. This oft-noted human ability is surely remarkable, but can it do explanatory work, and can it in turn be explained? Re-labelled 'memetics' this idea has generated excitement, but little insight. Sterelny's great achievement is to transcend the platitude, to provide an illuminating account of the phenomenon. Ironically, to make his case he has to overcome the trait itself, at work in the transmission of the human sciences. For the great advances made by Chomsky and his followers are now part of the developmental environment of all students of human thought. Chomsky's intellectual descendants have taken his approach to language as a paradigm, applying it to every facet of perception and cognition, proceeding via Fodor's modularity of mind to the massive modularity hypothesis championed by evolutionary psychologists. When a regime is overextended, revolution is in order. Just as Chomsky overturned Skinner's attempt to govern language with the (previously effective) apparatus of reward and punishment, Sterelny revolts against attempts to understand all human thought as analogous to the innate, modular language organ. *Thought in a Hostile World* is a detailed inventory of the tools needed to resist the dehumanising tyranny of modularity. And what exciting philosophy it is! Sterelny blends empirical detail with analytical rigour into an explosive mix, deploying his charges to recover two pieces of lost territory: folk biology and folk psychology. Neither, he argues, is like the language module, either in its operation or in its development; and he gives good evolutionary reasons why language should be an outlier, not an archetype.

#### Entrenched automated skills

What's the alternative? Sterelny argues persuasively that both our ability to interpret and predict others, and our facility with biological categories, are automated skills, rather like the ability to do maths, to play chess or to perform music. Crucially, they are not informationally encapsulated. Nor are they innate. Indeed, Sterelny delivers what many have only promised: an account that moves beyond the traditional nature-nurture dichotomy. Individuals acquire entrenched abilities through developing in an informationally rich environment. That developmental environment is copied down the generations in order to support the acquisition of entrenched skills, and is itself cumulatively improved to that end. This "cumulatively engineered epistemic environment" is only useful because learners are endowed with certain cognitive traits, principally the ability to imitate in a way which is sensitive to a demonstrator's goals and highly faithful to her technique. In hominid evolutionary history, epistemic environments and learning mechanisms have co-evolved, together constituting a rich "developmental scaffolding". According to the old picture, evolution gives us innate mechanisms whose development, although it might be triggered, is relatively informationally insensitive to the developmental environment. Sterelny's insight is that evolution has also given humans special learning mechanisms, whose purpose is to produce mature abilities in a way which is exquisitely sensitive to information in the developmental environment. We have evolved to be phenotypically plastic (resulting in humans' greatly extended developmental period). That makes the nature-nurture dichotomy particularly fallacious in our case.

Where information in the environment is stable over evolutionary time, the environment is "informationally transparent", so that the environmental information can be exploited by domain-specific encapsulated systems. But evolution itself tends to erode such informational links. Prey evolve to disguise signs of their presence, or to mimic other species which honestly signal their undesirability. Agents evolve away from stereotyped behaviour and start acting in ways that are sensitive to the competitors that are observing them; thus, they become less predictable to those observers. Organisms start to behave in ways that are designed to acquire or broadcast information. Such "translucent"

informational environments select for mechanisms that track important ecological features robustly, relying on multiple independent cues. The translucency is most marked in social species, where conspecifics' future actions are crucial ecological features. Fast and frugal heuristics can't cope in these conditions. They require simple cues, and can easily be exploited by competitors. So there is pressure towards flexibility and learnability.

Thus, Sterelny argues, much human behaviour is not produced by modules, whose development is strongly canalised and whose operation is informationally encapsulated. Especially when it comes to predicting and interpreting others (folk psychology), the problem is just not susceptible to a modular solution. Nor does a massively modular mind avoid the so-called frame problem – the problem of making global assessments of the significance of new information. A species of the frame problem persists for coordination amongst cognitive modules. Furthermore, a supposed folk psychology module would surely face its own mini frame problem within the large array of information processed by the module. Sterelny does not offer an answer to these puzzles, but at least his entrenched skills talk to each other. They are not informationally encapsulated so as to preclude a solution.

### Poverty of the Stimulus?

How, then, are entrenched skills acquired? Don't they face their own poverty of the stimulus problem? Sterelny narrows the poverty gap from both sides, explaining both why the environment is much richer than supposed and how individuals can use it efficiently. Starting with the latter, Sterelny notes the power of perceptual modules that are able to classify the environment functionally and so to keep track of salient ecological conditions (something like Gibsonian affordances). Applied to the interpretation of others, this is an important intermediate between behaviour reading and mind reading: the ability to group others' actions together functionally, rather than in terms of perceptual similarity. There is evidence that apes understand conspecifics' movements in terms of such behavioural programmes. Working from the other end, Sterelny highlights the support that can be provided by the learning environment. This is where his evolutionary understanding comes into its own. Only under very specific conditions can the learning environment itself be cumulatively modified over the generations so that it becomes part of a developmental scaffold, capable of supporting the acquisition of complex entrenched skills. Thus, mature entrenched skills owe little to innate knowledge, but are built by learning mechanisms, perceptual modules and an informationally rich developmental environment acting together; all three having been designed by co-evolution to produce the mature ability.

This insight builds upon behavioural ecologists' understanding of niche construction: cases where agents change the developmental environment of their offspring in a stable way. Something as simple as a parent's physical location can act as a stimulus enhancement, explaining why offspring get a greater opportunity to experiment with and learn about some types of food than others. Such factors alone can lead to stable differences between groups of organisms. At a second level of complexity, such group differences can be locked in as different elements of a suite of mutually reinforcing behaviours become stable. At this level, the culture of the group becomes a selective pressure in its own right. However, niche construction is most powerful when the effect of organisms on their niche can be passed on with enough fidelity to be progressively improved. Sterelny argues that humans provide the best (perhaps the only) example of such cumulative downstream epistemic engineering, allowing our evolution to accelerate beyond phylogenetic speeds.

At this third level social learning allows innovations, discovered in ontogeny, to be accumulated and stably transmitted to future generations (as if they had arisen in phylogeny). Sterelny calls this Tomasello's ratchet. Tomasello's ratchet only starts to turn when two conditions are satisfied. The first condition is cognitive, that organisms should be able to learn by high fidelity imitation of behavioural programmes. That allows information about technique to be transmitted and accumulated. Only in humans is high fidelity imitation pervasive (and also sensitive to the goal of the action demonstrated). The second condition is social: that organisms should live in highly cooperative groups. Only then can the advantages of sharing information outweigh the competitive edge that an organism gains from keeping its fitness-increasing information to itself. That is,

Tomasello's ratchet requires group selection, which Sterelny argues has been a particularly powerful force in recent human evolutionary history.

Every argument from poverty of the stimulus to the existence of an innate module must be re-examined in the light of these insights. Each could be an acquired entrenched skill. How much does perceptual sensitivity to functional categories simplify the learning problem? How rich is the developmental scaffolding, and how rich could it have been in the evolutionary past? Will selection have favoured a mechanism that is developmentally canalised or plastic? The case goes different ways in different domains. Comprehending and producing the syntax of natural language has been a stable feature of human evolution, not under competitive pressure, so is likely to have become canalised via a Baldwin effect. Intuitive understanding of the physics of objects and tools is similar. And in both cases we do indeed find a learning gap to be bridged by innate knowledge. However, the same considerations lead to the opposite conclusion in other cases: the culturally universal tendency for people to recognise that all organisms are members of species (folk biology), and especially when considering humans' skill in interpreting others – folk psychology.

### Folk Psychology

Sterelny argues that the evidence for the functional independence and informational encapsulation of folk psychological skills is not strong. Indeed, it is highly unlikely that we could interpret others using a limited subset of our total information whose relevance can be predicted in advance. If we did, our predictability would immediately be vulnerable to exploitation by conspecifics, given that social intelligence likely has Machiavellian as well as cooperative functions. Furthermore, there is a "wealth of the stimulus" argument against the postulation of an innate module. Sterelny brings to light the robust scaffolding that supports the development of our entrenched skills in interpretation: perceptual mechanisms tuned to functional categories of behaviour; the copious interactions a child has with agents who already interpret others; and the huge assistance offered by and through language, being taught, hearing interpretative stories and explicitly rehearsing interpretations.

Of course, once hominids acquired basic interpretative capacities, there would have been selective pressure towards developing better folk psychology. However, such pressure can lead to improvements in the developmental scaffolding rather than genetic assimilation. Amidst preoccupation with the Baldwin effect the former possibility is usually overlooked. Given the variability of human culture, folk psychology is a domain where there is evolutionary selection for phenotypic plasticity, as argued above, and the accumulation of information to improve interpretative skills is more likely found in cumulative downstream epistemic engineering of the developmental environment.

Since folk psychology performs such a useful function, surely its tenets must be true? Sterelny is unsatisfied with this simple move, so he has an independent argument that some human behaviour is caused by contentful beliefs and desires (Part I of the book). We really have beliefs if some actions are caused by states that register the world but are decoupled from any specific action. That is, the test is for breadth of behavioural response. The empirical evidence is inconclusive, but other animals may have decoupled registrations of space (maps), or of the causal properties of their environment. However, the strongest pressure to evolve belief states comes from social life. Humans, and perhaps some other primates, are able robustly to track some psychological causes of behaviour, and to deploy that information in a wide range of responses. These representations need to be decoupled because: (1) intentions are rarely clear from a stable set of cues; (2) the information is often needed later, rather than for immediate action; and (3) the information is usually relevant to more than one action. Such increasing social intelligence allows organisms to control their own actions in more complex ways, making them harder to predict, thus selecting for even stronger prediction engines. Sterelny argues that this feedback loop provides one clear route for the evolution of decoupled registrations. He is more sceptical about decoupled preferences. He worries that much human behaviour is caused by hedonic drives, rather than represented preferences. Only rarely do humans act on the range of well-ordered preferences that are ascribed by rational choice theory. In the end he accepts that some human behaviour, at least, is likely caused by decoupled representations of preference (ie, desires). He then expresses doubts about the nature of the contents of

beliefs and desires – an argument about which I will register a reservation.

## Two Reservations

Certain core commitments are shared by folk psychology and its refinement in sciences like cognitive psychology and rational choice theory – namely, that some actions are caused by an architecture of (i) beliefs (registrations of the world) and desires (preferences / ways the world could be); (ii) with contents; (iii) that we can identify by interpreting others. According to what Sterelny calls the Simple Coordination Thesis (SCT), these explanations work because the core commitments correctly describe the architecture of human minds. So far, so good. However, Sterelny saddles the SCT with a further commitment: that content is a matter of how an agent is connected to the external world, and further, that there should be a single connection property which is the content of all beliefs and desires (pp. 7 & 231).

Sterelny is surely right to doubt that there is such a univocal connection property to identify with content (pp. 17 & 233), and it is true that some philosophers have looked for one. However, the vindication of the core commitments of intentional psychology does not require that thought contents be some connection property (ie, that they be reducible to or identifiable with some connection property), let alone a single connection property. The SCT requires only that content can be naturalised. One important naturalistic theory of content – the teleosemantics of Millikan (1984) and Papineau (1993) – does not treat content as a connection property at all. Rather, the objects and properties (or conditions) mentioned in giving the content of a belief, say, enter into a different kind of explanatory relation with the thinker. Teleosemantics says that the content of a belief is the condition common to explanations of why behaviours prompted by that belief in the organism's evolutionary history were successful. So, belief contents are common explanatory conditions, and connection to the objects and properties mentioned in those conditions is not fundamental. That is a merit of teleosemantics, for it allows that an organism can represent something to which it is not causally sensitive. Teleosemantics is not the only way to go, but it does demonstrate that the SCT need not be committed to content being a connection property, let alone a single connection property. Sterelny is just right when he characterises content as that which explains the existence of cognitive states (p. 231). And of course, contents must 'pick out some real, natural relationship between minds and the world' (p. 232). But it is a non sequitur to conclude that naturalisation of content requires that there is a single connection property between an agent's cognitive states and environment (p. 233). Something like the SCT is in better shape than Sterelny acknowledges, even though naturalising content is more difficult than heretofore supposed.

My second reservation is a request for elaboration, without which Sterelny's theory cannot properly be empirically assessed. His alternative to modules for folk biology and folk psychology relies upon humans' capacity to take abilities learnt informationally (through empirical generalisation, etc.) and automatise them, so that they become entrenched skills which operate quickly, but are nevertheless informationally open. The examples of chess, maths and reading show that such entrenched skills can be acquired. But it would be useful to know much more about how the process of automatism works, and how the resulting skill operates in the mental economy. A greater understanding in these areas will further constrain the theory of cumulative downstream epistemic engineering, making it more detailed and more plausible. Furthermore, knowledge of the mechanisms of automatism will be important in assessing which mature abilities are in fact products of cumulative downstream epistemic engineering. That such data is sparse is not a criticism of Sterelny. Rather, his book does the invaluable job of showing that the need for further empirical work on automatised skills is much more pressing than was previously realised.

## Conclusion

Sterelny can be seen as harvesting the first fruits from a seed planted by Richard Dawkins in 1976 when he invented the word 'meme'. Sterelny hardly mentions the connection, distancing himself from much empty talk of 'memetics' and insisting that organisms, not artefacts, remain the vehicles of transmission. However, he vindicates Dawkins' insight

that, as well as the phylogenetic and ontogenetic timescales, some organisms are built over the cultural timescale. Sterelny has succeeded in explaining how a cognitive capacity can be built in cultural time. The entrenched domain-specific skills that result have some of the marks of modularity, but are nothing like innate. They are due to the evolution, in hominids, of a new inheritance channel. It is commonplace to notice that all traits arise due to a combination of nature and nurture. Sterelny's achievement is to say how. The surprising result is that, at least for psychological traits, it is incoherent to ask how much of the mature state is due to the environment, or the extent to which it is innate. Something social and environmental becomes something biological.

There have been many attempts to discern which of the many traits that are unique to humans (language, encephalisation, tool use, etc.) are responsible for our remarkable proliferation. Sterelny has uncovered a deeper fact. Hominids have been selected for phenotypic plasticity, and the resulting developmental scaffolding has allowed a new set of inheritance channels to be integrated into our development, so that the rate of flow of information between generations far exceeds that which can be carried by genes. These new inheritance channels furnish humans with an expanded space of evolutionary possibilities.

*Thought in a Hostile World* is a significant contribution to the understanding of human thinking. The progress it makes depends upon Sterelny's interdisciplinary approach, combining mastery of a range of empirical data with a philosophically motivated concern for deep theoretical issues. Philosophers are particularly well placed to make such advances, their logical training making them adept at building and scrutinising theories, and their freedom from the ties of the lab allowing them to work on a canvas that is broad enough to encompass many disciplines. Sadly, philosophers rarely get stuck into empirical work with Sterelny's courage and vigour. He has the ambition to contribute to debates that are internal to various natural and social sciences (for example, behavioural ecology and game theory), eschewing the safety of the most abstract philosophy. That makes his task harder and his work more open to criticism. But, as *Thought in a Hostile World* demonstrates, the effort can produce work of deep insight and wide significance.

Nicholas Shea  
Faculty of Philosophy  
University of Oxford  
6 April 2005

### Acknowledgements

Many thanks to Susan Hurley, Mohan Matthen, Hannah Pickard and Kim Sterelny for helpful discussions of the book.

### References

- Frank, R. (1988), *Passions Within Reason*. London / New York: Norton.  
Millikan, R. (1984), *Language, Thought and Other Biological Categories*. Cambridge, Mass: MIT Press.  
Papineau, D. (1993), *Philosophical Naturalism*. Oxford: Blackwell.