Australasian Journal of Philosophy
Vol. 82, No. 3, pp. 503–511; September 2004

SOME QUESTIONS FROM THE NOT-SO-HOSTILE WORLD[1]

Stephen Stich

Kim Sterelny has written a terrific book! It is brimming over with important and original ideas, rich in empirical detail, and written in a lucid and engaging style that makes it accessible to readers with a wide variety of backgrounds. The book does not fit comfortably into familiar categories since it makes significant contributions to philosophy, evolutionary biology, anthropology, and cognitive science. Sterelny addresses cutting edge issues in each of these disciplines with impressive sophistication and truly remarkable erudition. This is interdisciplinary work at its best.

One of the most valuable lessons of the book grows out of this broad interdisciplinary approach. Debates about many of the issues Sterelny discusses, including nativism, modularity, theory of mind, the evolution of cooperation, and the emergence of culture, are all too often carried out within the confines of relatively narrow traditional disciplines where the available options are very limited indeed. But, as Sterelny makes clear, once one begins looking beyond the boundaries of those disciplines, the range of options broadens in exciting ways. So, for example, for a generation, philosophers and cognitive scientists have been debating whether central components of language, folk biology, theory of mind, and a host of other human capacities are innate or acquired. But while these debates were grinding on, work in evolutionary biology and biological anthropology was making it increasingly clear that these are not the only options. Downstream niche construction—changing the environment in ways that affect future generations—can be a powerful non-genetic form of inheritance. And Sterelny argues persuasively that when niche construction is cumulative and modifies the *epistemic* environment, it can provide scaffolding that makes it possible to acquire skills and cognitive capacities that could not be otherwise acquired. When those capacities become widespread, they create a new environment in which natural selection may favour genes which modify cognitive capacities in other ways. Is the product of this process innate or acquired? After reading Sterelny's rich account of downstream, cumulative epistemic

[1] I'm grateful to Kent Bach, Peter Godfrey-Smith, and Shaun Nichols for their helpful advice.

niche construction, it is hard to take the question seriously. In his concluding chapter, Sterelny says rather modestly that he wants to put niche construction 'on the table as a candidate explanation of distinctively human cognitive capacities' [230]. But he has, I think, done much more than this. He has permanently altered the terrain on which battles over nativism and modularity will be fought.

It would be easy to continue on in this vein, recounting the many virtues of this extraordinary book. But author-meets-critic exchanges like this one are most useful, I think, when the critics raise questions about the author's views and arguments which the author can then address. I've got *lots* of questions, Kim, and in the limited space available, I'll raise some of them.

Staking out a position on the eliminativism debate is one of Sterelny's central projects in the volume, and while he has lots of interesting and original things to say about the issue, I am less clear than I would like to be both about what his view is, and about the arguments that get him there. As Sterelny sets up the debate, the terrain is marked by a pair of extreme positions. One of these, widely credited to Jerry Fodor, defends what Sterelny calls the 'Simple Coordination Thesis' which he unpacks as follows:

> (a) Our interpretative [= folk psychological] concepts constitute something like a theory of human cognitive organization; they are a putative description of the wiring-and-connection facts;[2] (b) Our interpretative skills depend on this theory, and our ability to deploy it on particular occasions; (c) We are often able to successfully explain or anticipate behaviour because this theory is largely true.
>
> [6]

At the other extreme are the Churchlands who 'argue that though the interpretation facts purport to describe the wiring-and-connection facts, they do a horrible job' [7]. Sterelny ends up someplace in the middle, offering

> a qualified and partial defense of the idea that folk psychology identifies some fundamental organizational features of the human mind. The folk have got something important right about how our minds work. But they have not got as much right as, for example, Jerry Fodor and Fred Dretske have supposed.
>
> [viii]

[2] The wiring-and-connection facts are 'facts about our internal organization (the wiring facts) and the facts about how that organization registers, reflects, or tracks external circumstances (the connection facts)' [4]. The terminology is due to Peter Godfrey-Smith.

But Sterelny's discussion of the issue makes relatively little contact with the extensive philosophical literature in which the merits of folk psychology have been debated (see, for example, Christensen and Turner [1993]; Stich and Warfield [1994], Ramsey [forthcoming]). As a result, when I reached the end of the book I found it hard to say where, exactly, Sterelny thinks that Fodor, Dretske and the folk are wrong, or why. What Sterelny thinks the folk have gotten right is much clearer; but even there there's a puzzle about how Sterelny's project meshes with the ongoing debate. I'll begin with this part of his account, since I can offer an interpretation of what he *might* be up to.

'Folk psychology', Sterelny maintains,

> has the following minimal commitment: each of the categories of belief and preference correspond at least roughly to organizational features of our cognitive architecture. We form and use *decoupled representations*[3] and we form and use representations of the *targets of our actions*. ... If nothing in human cognitive systems corresponds to beliefs and preferences, then folk psychology does not describe even the gross architecture of our cognitive system.
>
> [30]

At this point, Sterelny launches into an extended account of how a cognitive system exploiting the 'gross architecture' of decoupled representations and representations of the goals or targets of our actions might have evolved. It is a fascinating, wide ranging discussion, in the course of which Sterelny introduces a fist full of new theoretical notions, including transparent, translucent, and opaque environments, single-cued and multi-cued tracking, broad-banded and narrow-banded response breadth, and more, and makes a convincing case that these ideas earn their keep by clarifying important issues about the psychology and evolution of primate cognition. This is great stuff! But how does all of this contribute to the *eliminativism* debate? To be sure, had Sterelny argued that decoupled representations and preference-like representations of goals could *not* have evolved, *that* would pose a serious problem for the defenders of folk psychology.[4] But instead, he argues that in the hominid line these features of cognitive architecture could have, and probably *did* evolve. So the folk and their friends have nothing to worry about on this score. What is puzzling about all this is that, as far as I know, no critic of folk psychology has ever suggested that the folk were wrong

[3] 'Decoupled representations' are 'internal states that track aspects of our world, but which do not have the function of controlling particular behaviors' [29].

[4] At least for those of them who *believe* in evolution. From time to time, I suspect that Fodor doesn't.

about these very basic features of cognitive architecture.[5] Indeed, prior to Sterelny, no one had clearly drawn the distinction between decoupled representations and representations that have the function of controlling particular behaviours. So we didn't even have the conceptual tools to *raise* the issues of the existence or evolvability of decoupled representations.

'Where it doesn't itch', Quine once said, 'one ought not to scratch'. Could it be that Sterelny has spent three chapters of his book flouting Quine's wise advice? I'm inclined to think there may be a more charitable interpretation. Though he is hardly forthcoming on the issue, perhaps what Sterelny would say is that it *does* itch, and rather badly, though because they have not taken evolutionary questions seriously enough, neither friends nor foes of eliminativism have noticed. Or, to put the point more directly, perhaps Sterelny's idea is that, though none of the critics of folk psychology have noticed it, there *is* a serious issue about how decoupled representations and states representing targets of actions might have evolved. On my reading, then, Sterelny is both raising a new problem for the friends of the folk and offering a solution. So here's my first question, Kim. Is *that* what you had in mind?

Let me turn, now, to the other part of Sterelny's middle-of-the-road position. The folk, he argues, have got some important stuff right, but not as much as Fodor and Dretske suppose. Where, exactly, do the folk and their defenders go wrong? One way to approach this question would be to run through some of the widely discussed arguments aimed at showing that folk psychology is mistaken—arguments that Fodor, Dretske, and others have attempted to rebut—and to indicate which of them do in fact give us reasons to think that folk psychology is mistaken.[6] Does Sterelny think that any of these arguments give us good reasons to think that Fodor, Dretske, and others are too optimistic about folk psychology? Apparently not, since with a single exception that I'll come to shortly, Sterelny does not discuss *any* of the arguments against folk psychology that have dominated the eliminativism literature. What he does argue is that one argument that often encourages Fodorian optimism is weaker than many have thought.

> Defenders of the Simple Coordination Thesis are fond of the argument from success. The successful use of our interpretative concepts in ordinary day-to-day interactions shows that those concepts describe the cognitive architecture

[5] Sterelny is of no help here. He doesn't offer any references to eliminativists who challenge folk psychology on these issues, or to defenders of folk psychology who raise the issue as devil's advocates.

[6] For a catalogue of these arguments, see Stich [1996: 16–29].

> of our mind well. I shall argue that the power of that argument is much overstated.
>
> [90]

In setting out his critique of the argument from success, Sterelny has lots of interesting things to say about how predictive success might be achieved by strategies which *don't* invoke a largely true theory about the 'wiring-and-connection facts'. Rather, he maintains, 'a good deal of our predictive efficiency may rest on other cognitive adaptations for interpreting others' [228]. Though Sterelny does not claim that his case against the argument from success is conclusive,[7] let's assume that that he's right—that much of our success in predicting behaviour could be achieved without a largely true folk psychology. At most, what that shows is that one of the favourite arguments of those who defend the Simple Coordination Thesis won't work. But, of course, it does *not* show that the folk don't have and use a largely correct psychological theory, nor does it give us any indications where the folk go wrong. Moreover, as best I can tell, the rest of the book is of no help on this score. Though he asserts, quite clearly, that the folk haven't got as much right as Fodor and Dretske have supposed, *he never tells us what the folk have gotten wrong*. But perhaps I've missed something. Do help me out, here, Kim. You've told us what you think the folk have got right about how our minds work. What do you think they have they gotten *wrong*? And what are the arguments for your view?

I mentioned earlier that there is one eliminativist argument Sterelny does consider, and on Fodor's view it is the one that worries philosophers the most. 'The deepest motivation for intentional irrealism', Fodor tells us, 'derives not from ... relatively technical worries about individualism and holism ... but rather from a certain ontological intuition: that there is no place for intentional categories in a physicalistic view of the world; that the intentional can't be *naturalized*' [Fodor 1987: 97].[8] But here, again, I find Sterelny's view elusive. Indeed, since the project of 'naturalizing the intentional' has played such a large role in the work of Fodor, Dretske, and other defenders of folk psychology, it's a bit disquieting that, until one reaches the last ten pages of the book, it looks like Sterelny is simply going to ignore the issue. Throughout the book, the notions of representation and content are invoked freely, with no hint that worries about their naturalistic credentials

[7] 'How secure is the premise that folk psychology is basically true? ... The Simple Coordination Thesis depends heavily on an argument from success to truth. ... I do not reject this argument' [228].

[8] 'Intentional irrealism' is the view that there are no intentional states. If that's right, then if the folk think that beliefs and desires are intentional, they are very wrong about something very important.

may be the 'deepest motivation' for eliminativism. When Sterelny does get around to the topic, what he says is a bit puzzling. He tells us that he *used to believe* that there is a 'relatively straightforward *evolutionary vindication* of the idea that representational properties are natural and causally salient properties of cognitive states' [231]. But his view has changed. Now, apparently, he thinks that there is no 'single connection property'—indeed, no 'single natural relationship' with which representation or 'aboutness' can be identified [233]. Rather, there are lots of 'connection properties' that are 'evolutionarily significant' and he doubts that 'these connection properties are all species of a single genus' [ibid.]. 'Moreover, if the picture of the evolution of cognition sketched in this book is in the right ballpark, evolutionary considerations give us no reason to expect to find a single connection property onto which the folk have locked' [234]. This sounds like bad news for Fodor and for the folk. But in the next three sentences Sterelny tells us not to worry.

> The good news is that the compatibility of folk psychology with an integrated science of human cognition does not depend on aboutness (or truth) picking out a single natural relation. Folk biology would be not be catastrophically undermined if 'species' turned out to be ambiguous; if, for example, plant species turned out to be a different biological phenomenon from animal species. Likewise, we are not faced with a forced choice between eliminativism and the Simple Coordination Thesis.
>
> [234]

And with those reassuring words, Sterelny turns to other matters.

All of this, I'm afraid, goes by a little too quickly for a simple fellow like me. To explain my puzzlement, let me start with the last bit about there being no forced choice. Recall that the Simple Coordination Thesis claims that folk psychology constitutes a theory about the wiring-and-connection facts, and that the theory 'is largely true' [6]. Eliminativists like the Churchlands maintain that folk psychology does 'a horrible job' [7] at describing those facts; so the Churchlands think that folk psychology is largely false. How could it be that there is no forced choice between a theory being largely true and it being largely false? Perhaps what Sterelny is claiming is that the phenomenon of 'representational heterogeneity' indicates that folk psychology has indeed made an error of *moderate seriousness*—it's not one of the minor errors that Fodor happily acknowledges the folk are likely to have made, nor is it the sort of disastrous blooper made by the advocates of phlogiston and witches—favourite Churchland examples. So is that it, Kim? Are you telling us that representational heterogeneity indicates that folk have made a serious mistake, but not a 'catastrophic' one?

My own view is that the worry about 'naturalizing' the intentional, and much of the literature that attempts to carry out the project, is simply a muddle, since those who play this game have never given a clear account of what it would take to naturalize the intentional, or of why it would be a bad thing if it can't be done [Stich and Laurence 1994]. Though Sterelny does not address my worries—there's no reason he should—much of what he says about the 'naturalization project' [231] suggests that he thinks there is some reasonably clear account to be given about what it would take to naturalize meaning or content, and plausible arguments about what our reaction should be if it turns out that it can't be done. Is that what you think, Kim? If so, are you offering any hints? What is the 'naturalization project'? What would we have to do to naturalize intentionality (or aboutness or truth)? And if it turns out not to be possible, what conclusions should we draw?

I turn, now, to a cluster of questions on a very different topic. To get to them, I'll have to make a quick dash through some pretty difficult terrain. In Chapter 10, Sterelny sets out a rich and densely argued case against the massive modularity hypothesis that has played a central role in the work of leading evolutionary psychologists, including Leda Cosmides, John Tooby, Steven Pinker, and David Buss. Chomsky's work on language has been enormously influential in evolutionary psychology, despite his own scepticism about evolutionary explanations in psychology, and the Chomskian idea of a 'language organ' or module has been taken as a paradigm in explaining many other psychological capacities. Sterelny argues that this is a mistake because, though a modular account of important parts of language processing may well be correct, language is an 'outlier' [178] and some of the factors that facilitated the evolution of language do not apply in other domains. One of those factors, Sterelny argues, is that in the case of language (in contrast, say, with the case of sexual jealousy), 'there is no arms race between deceptive signaling and vigilant unmasking. ... Where there is no temptation to deceive, co-evolutionary interactions will tend to make the environment more transparent and the detection task less informationally demanding' [180]. And under those conditions an 'encapsulated mechanism' or module might well be able to handle the job. But wait, one might think, surely there is lots of deception in linguistic communication. Isn't that what spin doctors and press agents are for? Sterelny's response to this obvious objection is to distinguish two stages in language decoding: *understanding* and *acceptance*. The first stage requires identifying grammatical structure and communicative intention; in the second stage the listener must decide whether to believe what the speaker is trying to get her to believe. It is the understanding stage that 'operates in a social domain in which there is no danger of defection. It is in the interest of speakers to make the detection of syntactic structure and communicative intention as easy as possible and it is

in the audience's interest to recognize that structure and those intentions' [184]. In order for all this to hang together Sterelny must maintain that 'the proximate function of speech is to signal a communicative intention' [ibid.]. For that to be plausible, of course, it must also be the case that speakers *do* signal communicative intention, and that hearers detect it. In defence of the claim that speakers signal communicative intentions and hearers detect them, Sterelny tells us a fanciful tale about a conversation between Old Bear and Two Aardvarks (distant ancestors of ours, I assume), who are concerned about what Hairy Max (an ancestor of Kim's?) may be up to.

Now, finally, I can begin to pose my questions. As Sterelny notes, the story he tells is inspired by Grice, as is 'the idea that *understanding an utterance involves recognizing the speaker's communicative intention*' [181, emphasis added]. I've never been much impressed by the Gricean account, and it has certainly taken its share of criticism over the years [Davis 2002; Kemmerling 2001; Siebel 2003]. Most of that debate, like so much philosophy, takes place in an empirical vacuum, with arguments ultimately turning on philosophers 'intuitions' about what beliefs and intentions speakers must have. Sterelny, by refreshing contrast, is a philosopher whose work is richly informed by empirical findings. And since he is interested in the process of interpretation or 'mindreading', he is well acquainted with the experimental literature that has been widely interpreted as indicating that young children and people with autism don't have beliefs *about beliefs*.

> In this developmental literature, the key idea is that a child does not really understand belief (and other intentional concepts) until she understands that others have and act on beliefs unlike hers. False belief tasks test for this ability. In one version of the false belief test, a child watches two puppets interacting in a room. One ('Sally-Anne') puts a toy in a box and then leaves the room. While Sally-Anne is out of the room, the other puppet moves the toy from the box to a drawer. Sally-Anne returns to the room, and the child onlooker is asked where Sally-Anne will look for her toy. Three year olds regularly predict that she will look where the toy now is, namely the drawer. Sometime between four and five, children predict that she will look in the box: they understand that Sally-Anne has a false belief and will act on it.
>
> [212][9]

In similar experiments, children with autism give the same responses as 3-year olds, even though their mental age, on standard IQ tests, is much higher.

[9] Sterelny makes an uncharacteristic (and unimportant) factual error in this passage. In the standard accounts, there is no puppet named 'Sally-Anne'; one puppet is named 'Sally' and the other is 'Anne'.

Now at first blush these findings pose a serious problem for the Gricean idea that 'understanding an utterance involves recognizing the speaker's communicative intention' since if three year olds and people with autism don't have beliefs about beliefs, they surely don't have beliefs *about other people's intentions to get them to believe that p by recognizing their intention to* . . . [please fill in your favourite version of the Gricean account]. It follows that three year olds, and autistic individuals with a much higher mental, age *don't understand utterances*. And, having had lots of extended and quite informative conversations with autistic people, and lots of extended though less informative conversations with three year olds, that strikes me as a *reductio* of the Gricean account of understanding an utterance. So here are my last two questions for you, Kim. Do you agree that the empirical literature poses a serious problem for Gricean accounts of meaning and understanding? And if so, how much damage does this do to your argument against massive modularity?

Rutgers University

REFERENCES

Christensen, S. and D. Turner, eds. 1993. *Folk Psychology and the Philosophy of Mind*, Hillsdale NJ: Lawrence Erlbaum Associates.

Davis, W. 2002. *Meaning, Expression and Thought*, Cambridge: Cambridge University Press.

Fodor, J. 1987. *Psychosemantics*, Cambridge: MIT Press.

Kemmerling, A. 2001. Gricy Actions, in *Paul Grice's Heritage*, ed. G. Cosenza, Turnhout: Brepols: 69–95.

Ramsey, W. forthcoming. Eliminative Materialism, *The Stanford Encyclopedia of Philosophy* Fall 2003 edn, ed. Edward N. Zalta URL: ⟨http://plato.stanford.edu/archives/fall2003/entries/materialism-eliminative/⟩.

Siebel, M. 2003. Illocutionary Acts and Attitude Expression, *Linguistics and Philosophy* 26: 351–66.

Stich, S. and S. Laurence 1994. Intentionality and Naturalism, *Midwest Studies in Philosophy*, 19: *Naturalism*, ed. by Peter A. French, Theodore E. Uehling, Jr., NotreDame: University of Notre Dame Press: 159–82.

Stich, S. and T. Warfield, eds. 1994. *Mental Representation*, Oxford: Blackwell.

Stich, S. 1996. *Deconstructing the Mind*, New York: Oxford University Press.