

The Return of the *Tabula Rasa*

STEFAN LINQUIST

University of Queensland

ALEX ROSENBERG

Duke University

*Thought in a Hostile World*¹ has four ostensible aims:

...[1] to develop and vindicate a set of analytical tools for thinking about cognition and its evolution... [2] to develop a substantive theory of the evolution of human uniqueness... [3] to explore, from this evolutionary perspective, the relationship between folk psychology and an integrated scientific conception of human cognition... [4] to develop a critique of, and an alternative to, nativist, modular versions of evolutionary psychology (p. viii).

Of these four aims, the most immediately interesting must be [2] and [4], a narrative account of what special features in hominid evolution led to our psychologically unique capacities, and the negative implications of that account for nativist theories that treat these psychological capacities as innate and genetically grounded. These two aims are interesting in part because their conjunction is to say the least, unexpected! *Thought in a hostile world* is nothing less than a Darwinian argument for something approaching Locke's *Tabula Rasa*.

The other two aims are less immediately interesting because the analytical tools Sterelny crafts are extensions of ones originally crafted by other philosophers, behavioral ecologists and social theorists (we

¹ Kim Sterelny, *Thought in a Hostile World: The Evolution of Human Cognition* (Malden, Oxford and Victoria: Blackwell Publishing, 2003).

describe his framework briefly below); and the conclusions to which Sterelny comes about folk psychology are relatively weak.²

Thus, in what follows we focus primarily on Sterelny's account of how our uniquely human psychological attributes evolved, and whether this account provides a superior alternative to the massive modularity theory defended by most evolutionary psychologists. Our general conclusion is that Sterelny does a remarkable job of developing a viable alternative to the massive modularity thesis, and that this achievement on its own marks an important development in evolutionary approaches to the mind. However, we argue that it remains unclear whether this alternative is in fact superior to at least some version of the view that the mind contains a system of cognitive modules. Our discussion concludes with some suggestions as to how further evidence coming from the fields of neuroscience and genomics might decide the issue once and for all.

Sterelny's conceptual toolkit is a montage of ideas drawn from behavioural ecology, philosophy and other related disciplines. Aside from assigning these concepts new and often catchier titles, Sterelny manages to outline a rough set of criteria for evaluating a system's cognitive complexity. This framework enables him to state fairly explicitly what makes humans psychologically unique, thereby identifying

² Aim 3 is an attempt to bring the implications of his evolutionary story to bear on what Sterelny calls the Simple Coordination Thesis—i.e. the thesis that folk psychology paints a more-or-less accurate picture of the “wiring and connection facts” at some deep level of human neurological organization. Sterelny explains that,

One way of evaluating [the Simple Coordination Thesis] is to place the thesis in an evolutionary context. That is my strategy, despite the fact that the Simple Coordination Thesis is not itself an evolutionary hypothesis. It is primarily a hypothesis about the proximal mechanisms of human action. Nevertheless, evolutionary considerations can be part of a total package of evidence in favor of a proximate hypothesis (p. 7).

The success of this project depends on how convincing an evolutionary story Sterelny can develop about the adaptive function of folk psychology, and how well that story can be supported by the available evidence about our actual evolutionary history. In what follows we consider in some detail just how convincing this story is, and what sorts of additional evidence might be required to support it further. But even with a plausible evolutionary story on the table it is unlikely that such considerations could sway the longstanding debate over the accuracy of folk psychology. As Sterelny himself notes, the Simple Coordination Thesis is a hypothesis about the structural aspects of the mind, whereas the sorts of evolutionary considerations with which Sterelny is primarily concerned (adaptationist inferences to the best explanation) describe the mind at the level of ultimate (Wrightian) functions. Since adaptive functions can be realized by a multitude of proximate structures, inferences from adaptive function to proximate structure are fairly inconclusive. So, however convincing a case Sterelny is able to prosecute either for or against the adaptive function of folk psychology, we doubt that it will hold much sway with a staunch eliminativist or, for that matter, with Fodor's Granny.

the explanatory target for an account of how the mind evolved. Humans are unique, Sterelny explains, in their capacity for generating “decoupled” representations. These are intentional mental states that can generate a broad range of behavioral outputs. According to Sterelny, we find the precursors for this type of representational system elsewhere in the animal kingdom. Creatures with “simple tracking systems” recognize important features of their environments by a single cue (think of the “fleebee detector” in frogs) and they typically respond to those cues with a stereotyped output. Slightly more sophisticated are the “robust” tracking systems that register features of their environment with multiple cues. The honey bee that locates nectar using petal-color, odor, and flower shape is a robust tracker on Sterelny’s account. However, like frogs, honeybees display limited behavioral flexibility—the identification of a nectar source releases a fairly stereotyped, fixed-action response. This is an important point for Sterelny: a cognitive system becomes a candidate for the ascription of a belief-like state (a decoupled representation) only when that representation potentially generates a broad range of behavioral outputs. Sterelny applies a similar analysis to motivational states. Unlike most creatures who come equipped with a fixed set of drives, humans can update their individual utility-functions in light of newly acquired information. In Sterelny’s parlance, humans have a “preference-based” as opposed to a “drive-based” motivation system. Thus, the human mind is flexible in at least two respects: it contains representations (putative beliefs) that are triggered by multiple inputs and which can give rise to a broad range of outputs, and it consists of motivational states (putative desires) that are capable of modification within an individual’s lifetime. Although this characterization of human cognition is not novel,³ Sterelny’s framework provides the resources for understanding how the mind might have arisen via a Darwinian process of gradual selection acting on relatively simple systems (as opposed, say, to the view that the mind arose *de novo* as a hopeful monster).

For instance, Sterelny suggests that representational systems tend to evolve in the direction of increased flexibility when they encounter “epistemically translucent” environments. An environment is epistemically *transparent* to some organism just when the features it needs to track (mates, prey, predators, etc.) are reliably identified by a single cue. For many visual predators, bright coloration is a reliable cue that a potential prey item is toxic. However, as soon as mimics appear in

³ For example, Jonathan Bennett develops a similar account of intentionality in *Linguistic Behavior* (1976).

the environment the epistemic situation becomes *translucent*. Under such conditions it might benefit the predator to start paying attention to additional cues—smell, shape, behavior, etc.—to distinguish models from mimics. This example illustrates how epistemically translucent environments tend to select for more robust tracking mechanisms. The example also illustrates how “biological hostility” (conflicts of interest that arise among species or between conspecific individuals) tends to degrade the epistemic environment. Social environments are especially hostile, Sterelny argues, because they contain agents who have a stake in being deceptive. Some social environments also contain agents who are capable of anticipating our behavior and acting in ways that will block or thwart those actions. Under these conditions an organism with inflexible behaviours would be easily exploited. Thus, Sterelny reasons, hostile environments tend to select for decoupled representations. And as an organism’s behavioral repertoire broadens it also becomes advantageous to evolve a more flexible motivational system for modifying—in one’s lifetime—the ways that those behaviours are prioritized.

Not only do these considerations provide a sketch of the selective pressures that might have led to the evolution of propositional attitudes in humans, they potentially explain another of our unique cognitive abilities: the capacity for generating meta-representations or a “theory of mind.” According to the Social Intelligence Hypothesis, humans evolved this ability to track their cohorts’ mental states in order to compete more effectively in a hostile social arena. However, Sterelny objects that the Social Intelligence Hypothesis is incomplete. Other species of primate live in social environments that are, arguably, no less hostile than the ones our earliest ancestors inhabited. Yet, these organisms do not exhibit the degree of cognitive flexibility one finds in humans. Nor is there good evidence that other primates (even chimpanzees) possess a theory of mind. A brief review of the chimpanzee “mind reading” literature reveals that their competence is patchy: on some other-directed tasks they do rather well, on others they are quite hopeless. These findings suggest to Sterelny that the underlying cognitive machinery chimps use to navigate their social worlds consists of a mosaic of relatively simple subsystems. Special-purpose *perceptual* (but not cognitive) modules are tuned to socially significant cues like gaze direction, facial expressions of emotion, and bodily gestures. The outputs from these modules feed into a more *general-purpose* pattern recognition mechanism that enables chimpanzees to identify correlations between social cues and behavioral outcomes. Sterelny argues that this apparatus would achieve the level of social competence one finds in other primates:

Agents who understand their social environment, and the behavioral regularities such environments enforce, are often in a position to anticipate the actions of other agents... Moreover, if agents can recognize the function of one action, they are often able to predict the next. Identifying a sound as a warning call, or a gesture as a threat, is predicatively salient. Identifying function is cognitively sophisticated. But it does not require a theory of mind, or even the ability to track the mental states of others. (p. 59)

This line of argument raises several important questions,⁴ but let us suppose that Sterelny is correct in his analysis of chimp minds. If other primates navigate hostile social environments quite successfully without fully decoupled representations and a theory of mind, then social hostility alone does not explain our unique cognitive architecture—something else must have driven the evolution of the energetically expensive brain tissue that supports these capacities in humans.

This brings us to Sterelny's own account of how the unique aspects of the human mind evolved. He identifies three factors in particular that were instrumental in the evolution of human cognition: *enhanced cooperation, cumulative niche construction and phenotypic plasticity*. These factors are thought to be mutually reinforcing. In order for humans to have become so cooperative, Sterelny reasons, selection must have operated at the level of the group. But group selection cannot take hold in a social system unless it has developed some means for solving the free rider problem. Cumulative niche construction—the cultural modification and transmission of traditions—facilitates this process by allowing for the rapid development and enforcement of punishment strategies. Culture further contributes to group selection by promoting phenotypic differences among groups while suppressing within-group differences. However, Sterelny argues that cultural transmission itself requires group selection to get up and running. Unless individuals receive a fitness payoff for cooperating, there is no incentive to teach a neat trick—e.g., the construction of composite tools—to any but immediate offspring. Finally, increased plasticity in brain development allowed humans to become more adept at social learning

⁴ For instance, Sterelny's account of chimpanzee social competence presupposes a sharp distinction between perception and cognition. Instead of using a conceptual (i.e. cognitive) mechanism for social navigation—a theory of mind—chimps rely on “low-level” perceptual machinery to categorize the social world into appropriate chunks. This is a lot to ask of a “dumb” module. How does the module do it? For instance, does the module possess something like an innate concept which it uses to identify patterns of similar behavior. If so, then it is not clear that Sterelny has provided an alternative to the nativist position so much as he has pushed the problem down to the level of nativist perception-modules.

and imitation, which were required for the reliable transmission of culture. Sterelny goes on to argue that these three factors would have favored the degree of flexible cognition we find uniquely in humans as well as a theoretical framework (folk psychology) for robustly tracking those mental states. His arguments take the form of a task analysis. By considering what sort of cognitive architecture is necessary for these capacities, Sterelny concludes that (a) a modular architecture would not have been up to the tasks at hand and (b) a more flexible cognitive system is required to generate these processes. In what follows we shall examine each of the three evolutionary processes Sterelny identifies, considering, in each case, whether his task analysis is sufficiently well developed to support these conclusions.

Sterelny notes that the extent to which humans are cooperative (especially in comparison to most other species) is a pervasive, yet under-appreciated fact about our social lives. Not only do humans hunt in groups using a distribution of labor, but there is also cooperation among the sexes in the rearing of children. The sharing of information about foraging strategies, tool construction and use also provides benefits to unrelated con-specifics. But, perhaps most importantly, humans engage in “moralistic punishment”, collectively punishing free riders who violate group-beneficial social norms. Once such a system is in place, members of a social unit can withhold benefits from a cheater or actively expel one from the group at relatively low risk to themselves while imposing, collectively, a high cost on the defector. Following Sober and Wilson (1998), Sterelny endorses a group-selectionist account of how moralistic punishment was maintained among ancestral humans. However, Sterelny rightly notes that moralistic punishment does not serve as an adequate explanation for the *origin* of these cooperative tendencies. We do not find sustained cooperative alliances in any of our living primate cousins, and one cannot assume that these tendencies were present in the earliest human societies. Thus, the appeal to moralistic punishment as an explanation for the origin of cooperation presupposes the very suite of psychological dispositions it sets out to explain.

A more satisfactory explanation for the origin of human cooperation is developed by Bingham (1999, 2000): all it requires is the emergence of stone throwing and clubbing (for which there is pretty good anatomical evidence in the fossil record). Throwing weapons enable all members of a coalition to simultaneously attack a more powerful bully while spreading and thinning his throwing targets. Sterelny identifies this as the “right sort of hypothesis” for the emergence of cooperation because Bingham does not ascribe overly sophisticated cognitive abilities to his club-wielding hominids. However, Sterelny criticizes

Bingham for other reasons. He points out that the use of weaponry could just as easily have made coalitions harder to maintain! A single alpha-male hominid armed with such a weapon can destroy incipient coalitions before they harden, nor is a bully likely to delay an attack until a coalition is in place. Sterelny also points to examples in the chimpanzee literature of lone individuals being killed by marauding bands. Since chimpanzees execute this gruesome task without the use of weaponry, Bingham is just mistaken that the development of weaponry was necessary in order to make coalitional enforcement cheap.

The scenario that Sterelny prefers draws upon Frank's (1988) theory of emotions as solutions to a commitment problem. On this account, motivating emotions (and their display) emerged first to signal the credibility of an individual's retaliatory threats even when the retaliations are more costly to the individual than desisting. Such signals could have emerged initially for their role in promoting inclusive fitness, for instance, in the context of protecting mates and offspring from iterated threats by other individuals. Later, as early human societies increased in size, Frank's commitment-securing motivations might have been applied to enforce cooperative alliances among non-kin—thereby making these social units more competitive at the group level. As Sterelny explains: “These capacities *could have been* recruited to further social purposes in environments with strong selection for cooperation. In other words, evolution *would not have to* build a Frank-like commitment mechanism from scratch in early hominids” (p. 141, italics added). Sterelny prefers this scenario to Bingham's because, “Frank's model does not presuppose explicitly formulated norms, moralistic punishment, or other cooperation amplifiers that kick in only once a cooperative social world has evolved” (p. 142). Among these amplifiers are language, scaffolded learning, information sharing, and social control by gossip and teasing, which group selection would have initially made possible.

Now, there are at least three things that stand out about Sterelny's analysis of the role cooperation played in structuring human cognition. Firstly, as the italicized passages (above) indicate, Sterelny's favored hypothesis is merely an account of how cooperation *might* have evolved in early human societies. The absence of independent evidence makes these debates inconclusive and reflects itself in the stigma of “just so stories” applied to incompatible adaptationist hypotheses. We revisit this issue below, offering some suggestions for the sorts of evidence that might settle the debate between Sterelny's hypothesis and its rivals. Secondly, there is a tension between Sterelny's emphasis on group selection and his more general claim that our ancestors' social environments were biologically hostile. When selection is acting at the

group level, members of the same social unit come to resemble a super-organism in their shared interest of out-competing rival groups. However, if group members have a common biological goal—cooperation—then the within-group epistemic environment should become increasingly transparent. This is important because, on Sterelny's model of cognitive evolution, epistemically transparent environments do not select for decoupled representations. On the contrary, transparent environments are thought to select for single-cue detection systems. Thus, if group selection was the driving force behind human cognitive evolution, this makes our remarkable degree of cognitive flexibility an even greater mystery.

We do not think that this problem is insurmountable for Sterelny. One possibility is that the relevant biological hostility that drove the evolution of decoupled representations occurred in the context of *inter*-group social interactions. For instance, the archeological record reveals that trade was an important feature of early human societies, perhaps so important that it was the key adaptation that enabled us to out-compete Neanderthals who were otherwise morphologically even better adapted to living in variable ecological conditions (Horan *et al.*, 2005). And the negotiations involved in trading between groups would have given rise to just the sorts of hostile conflicts of interest that are thought to favor flexible cognition. This alternative scenario raises a host of interesting questions. For example, does our folk theory of mind differentiate between the beliefs and desires ascribed to out-group vs. in-group members, and, if so, how well do these theoretical constructs track the mental states of individuals who fall into either of these two categories? We find it somewhat surprising that Sterelny's task analysis does not investigate the cognitive prerequisites for inter-group social navigation, especially given the emphasis he places on group selection.

This brings us to our third criticism of his task analysis. Sterelny's general hypothesis is that the emergence and maintenance of cooperation would have required a flexible or non-modular cognitive architecture. Yet, at the same time, his favorite explanation for the emergence of cooperation appeals to Frank-style commitment devices. On this view, it was our emotional reactions to free riders that supplied the requisite motivation to overlook the individual costs associated with punishment. The interesting thing to point out here is that this explanation presupposes a highly *modular* model of the mind. In order for a commitment device to function in the way Sterelny imagines it must be *encapsulated* from an agent's short-term calculations of the potential cost of punishing free riders, and encapsulation is the defining feature of a mental module (Fodor, 1983). So, it appears that the emergence of

cooperative tendencies required the evolution of a mental module that was not previously present in our less-cooperative ancestors.

One possible way of reconciling this tension is to posit three discrete steps in the evolution of human cooperation. Step one involved the evolution of a Frank-style module that motivated early humans to punish free riders while overriding utility calculations of the short-term costs of upholding such commitments. Then, once stable coalitions were in place, more sophisticated enforcement mechanisms (like moralistic punishment, gossip, teasing and the like) took over as the primary means for sustaining cooperation. These more sophisticated tasks would have (arguably) required a flexible cognitive architecture. Thus, step three would have involved a decreased dependence on the Frank-module, and the concomitant evolution towards increased flexibility. However, this is a rather unparsimonious suggestion. It is the move from step 2 to step 3 that we find particularly troubling. After all, why would selection favor the emergence of more sophisticated cognitive mechanism for enforcing cooperation when there was already a perfectly good module in place for performing this function?

Sterelny supplies an argument that can be recast as a defense of the three-step model. As social system becomes more cooperative, he suggests, the various pieces of information that an individual must process become more tightly integrated. And, Sterelny maintains that a massively modular architecture would perform poorly on this task:

A module is a more or less autonomous cognitive machine specialized for driving action in a specific domain. Theory of mind and social intelligence; technical intelligence and tool making; natural history are all candidate domains. In my view, the cooperation explosion casts doubt on the existence of independent domains of this kind. For example, technical skills are acquired via social intelligence. Imitation involves understanding the function of the actions of the model. Identifying an agent's goal when the agent is engaged in (say) a resource extraction task requires both social intelligence (reading the other agent) and foraging intelligence... If the division of labor is a deep feature of hominid history; if complex coordination is a deep feature of hominid history; if the iterated socially mediated learning of technical and natural history competences a deep feature of hominid history, then technical and natural history domains are not independent of the social domain, and vice versa (p. 144–145).

This is an interesting line of argument. However, the strength of Sterelny's proposal depends on whether (a) it is possible to imagine an alternative, modular architecture that achieves the requisite degree of domain integration, and (b) whether this architecture is equally or

perhaps an even more plausible candidate than the domain general mechanism Sterelny posits. We suspect that both conditions can be satisfied. The first thing to point out is that Sterelny is equating modularity with domain specificity: two modules are distinct, on this view, if they process different sorts of information (social rank vs. foraging strategies vs. tool making, etc.). However, modularity is a thesis about the architecture of the mind. Module A is distinct from module B if A has access to some “proprietary database” that B cannot access, or vice versa (Fodor, 1983; Samuels, 2000). The important thing to realize is that the same piece of information can be stored simultaneously in multiple databases. In other words, the Massive Modularity Hypothesis is not committed to the (rather outlandish) idea that all of the information about foraging is stored in one database while all of the information about social rank (say) is stored somewhere else.⁵ Mother Nature can construct modules however she chooses. If information about social rank is relevant to foraging decisions, then it makes sense to store these two pieces of information in the same database. At the same time, information about social rank might also be stored in a database that contains information about tool making. What this debate comes down to is whether all types of information are equally relevant to all others—i.e. whether hominid social environments are “Quinean” in Fodor’s (1983) parlance. Only then would we expect selection to favor the sort of domain general architecture Sterelny posits. As it happens, we don’t think that early hominid social environments were epistemically interwoven to the degree that Sterelny suggests. Granted that information about social rank is relevant to many of the tasks early humans were engaged in. There are many other tasks that are less obviously interconnected. For example, information about foraging strategies has apparently little to do with the decisions governing mate choice (Cosmides & Tooby, 1994). So, it seems plausible that social dominance information is distributed in the human brain across a wide array of databases. But it is unlikely that human cooperation would have required a domain general processor capable of integrating information across any potential set of domains whatsoever. Indeed, an encapsulated mechanism seems necessary at least for the origin of cooperation.

⁵ Here is another way to state our objection: the term “domain” is ambiguous. The external sense of “domain” or (e-domain) refers to a set of actual or potential states of affairs in the world that have certain important features in common (all states of affairs that involve foraging, for example). The internal sense of domain (or i-domain) refers to the way that those states of affairs are represented in the head. Sterelny is equivocating between these two senses. That is, he fallaciously infers that a difference in e-domain implies a difference in i-domain.

Let us turn, then, to the second factor Sterelny identifies in his evolutionary story. Cumulative niche construction is the carving out of an environment that persists after the demise of the carvers-out (and so it is not just an extended phenotype). The niches that our ancestors and their descendents constructed and persistently improved were largely epistemic—strategies for engineering the cognitive development of our offspring so as to make adaptationally important features of the environment more salient. Here the ability to imitate become critical. It is an ability in which we exceed our primate cousins by many orders of magnitude. Sterelny coins the term “scaffolded learning” to describe the integration of teaching and imitating which, he argues, enable our domain general processing capacities to produce increasingly specialized outputs, so specialized that they look like they may rely on innate modules even when they do not.

But how might cumulative niche construction have selected for decoupled representations and contributed to the emergence of our unique cognitive capacities? Sterelny suggests that the progressive development of tools, hunting, and other technologies enabled early humans to inhabit an increasingly broader range of ecological environments. Ecological variability is another form of biological hostility that potentially selects for more flexible cognition (Godfrey-Smith, 1996). For instance, learning how to modify one’s foraging techniques to suit the local microenvironment, or determining how to fashion tools from locally available materials might have required a degree of cognitive flexibility achieved only with decoupled representations. On this view, it was adaptation to the physical rather than the social environment that drove the evolution of flexible cognition. Sterelny further argues that keeping up with rapid technological developments emerging from within a cultural tradition would have put a premium on cognitive versatility and, in so doing, supplied a further selection pressure for decoupled representations.

It is hard to deny that the appearance of culture and the rapid expansion of early humans across the globe would have been an important source of ecological variation, and niche construction would have certainly facilitated this variation by human constructions. However, there is another side to the niche construction coin. As Sterelny himself notes, niche construction is also a means by which many organisms *buffer* themselves against environmental fluctuations. Termite mounds—a now classic example of this phenomenon—are structured so as to protect the organisms that inhabit them from extreme fluctuations in temperature. Similarly, the domestication of animals and plants meant that humans were able to bring food sources with them as they migrated. Parenting strategies, marriage practices, religious rituals and

other cultural phenomena are also portable strategies for epistemic engineering that are potentially unaffected by geography. Thus, although niche construction allows humans to occupy a more diverse range of *physical* environments, it seems that in many cases it serves to stabilize the *selective* environment (*sensu* Brandon, 1990). Moreover, Sterelny's suggestion that technological advances posed a significant source of ecological variability might be overstated. It is only relatively recently, with the advent of agriculture that human technology has really begun to take off. The stone toolkits of the upper Paleolithic, though much more sophisticated than the hand axes that persisted for millions of years before, do not change so rapidly as to demand a specialized cognitive architecture to keep up. Thus, while niche construction increases variability in some aspects of the environment, it causes other factors to stabilize. If we assume that selection tends to favor a more hardwired or modular response to those features of the environment that remain stable (because a dedicated response is more economical to produce and less prone to developmental perturbation than one that relies on social learning), then the upshot is that niche construction ought to have selected for some hardwired cognitive modules alongside a general purpose learning device. And this, of course, is a departure from the model that Sterelny endorses.

Perhaps the third cognitive ability that Sterelny identifies as a uniquely human trait—our remarkable degree of plasticity—can settle the issue of whether (or the extent to which) the architecture of the mind is modular. Although it is widely agreed that humans are capable of acquiring a broad range of skills not selected for in the ancestral environment (e.g. chess, driving a car, mathematics, etc.), there is disagreement about the sort of underlying architecture required to support these capacities. On the massive modularity model, plasticity is achieved through modifications made to innately endowed algorithms. On the model Sterelny endorses, these skills are transmitted culturally and acquired *de-novo* during a prolonged period of cognitive development. The way that one goes about investigating the mind will depend on which model is endorsed. If the modular account is true, then the project is to first identify what the set of innately endowed algorithms are, and then to determine how they get modified through social and individual learning. On the alternative picture the first step can be dispensed with. If Sterelny is correct, the important cognitive adaptations are to be found “outside the head” (so to speak) in the form of socially transmitted learning strategies that enable a domain general processor to specialize on ecologically significant tasks.

Why think that the human mind is so plastic that a research project that attempts to identify innate algorithms would lead down a blind

alley? Sterelny points to three examples of cognitive plasticity: (1) our remarkable capacity to acquire and retain a diversity of skills, (2) developmental sensitivity in affect and evaluation, and (3) neural plasticity. His argument is that the degree of plasticity exhibited in all three cases could not be achieved with a highly constrained, modular architecture. Moreover, Sterelny contends that the degree of plasticity humans exhibit is itself an adaptation to a cooperative social environment that contains numerous opportunities for scaffolded learning. As we shall now argue, it is the first leg of this argument that is somewhat problematic.

Sterelny is impressed by the degree of variability in cognitive specializations or skills exhibited among human individuals and cultures. He thinks that these differences would not appear if the mind were constrained by a multitude of innately specified modules. However, modularity theorists will argue that those skills could only be achieved through the modification of existing algorithms, because a domain general processor would run up against the frame problem. Notice that in this debate everyone agrees that learning and culture are important, and almost no one thinks that there are no innate algorithms that structure these learning processes. Thus, the debate comes down to a matter of emphasis: how much of our broad skill-base is innately determined and how much is socially scaffolded? It would be most impressive, we think, if the anti-nativist could demonstrate that some trait which ought to be genetically specified in fact relies on a considerable amount of socially transmitted information. Sterelny cites some evidence for thinking that this is true of mate choice (Gangestad & Simpson, 2000; Fletcher & Stenswick, 2004; Simpson & Orina, 2004; Gray *et al.*, 2004). However, in one of the most thorough investigations on this topic conducted thus far, Schmitt (2005) found dramatic differences between men and women in their approaches to sex: “it can be concluded from these results that sex differences in sociosexuality are a cultural universal, supporting the basic tenets of parental investment theory... Culture has an important influence on sociosexuality, but biological sex is the larger and stronger predictor of human mating strategies” (Schmitt, 2005, p. 265). Thus, it appears that the verdict is still out on whether (or to what extent) human mating patterns are culturally universal, and, moreover, what this might mean for the proposal that these functions are executed by a modular mechanism. A similar objection applies to Sterelny’s second example, individual differences in affect. Sterelny cites variability in the disgust response as an example to support his anti-nativist model the mind. Humans, like many other creatures, will form a rapid and long lasting aversion to foods that have made them ill. However, the machinery that underlies this

capacity presumably involves a specialized learning algorithm (Rozin & Fallon, 1987). In fact, Garcia's now classic work on taste aversion learning in rats is frequently heralded as a paradigm example of an innate learning bias. Thus, it is somewhat odd that Sterelny should choose this as an example to support his anti-nativist position. The third line of evidence Sterelny cites to illustrate the magnitude of human cognitive flexibility comes from developmental neuroscience. Like most mammals, humans undergo a stage of significant neural growth early in development followed by a major die-back event (apoptosis) where various neural connections are pruned. There has been some suggestion that this is a developmental strategy where the organism relies on its interaction with the environment to selectively prune the neural connections that are unlikely to be needed in a given ecological context. Edelman (1987) refers to this process as "neural Darwinism" and Plotkin (1994) is similarly motivated to call the brain a "Darwin Machine." The point that Sterelny takes from this phenomenon is that two individuals could end up with quite different neural architectures through this developmental process as a result of having encountered different environments. Now, while we agree with Sterelny that, "it is unlikely that all of these neural variations are cognitively epiphenomenal" (p. 165), it does not follow that neural diversity or even neural Darwinism undermines the massive modularity thesis. One possibility is that neural plasticity supports massive modularity by allocating and occasionally reallocating domain specific, encapsulated functions to disparate brain regions. For instance, it has been demonstrated that when ferret's nerves are disconnected from the visual cortex at birth and reconnected to the auditory cortex, they are still able to see (von Melchner *et al.*, 2000). This is a particularly striking result in light of the fact that exponents of modularity such as Fodor have identified the visual system as a paradigm case of a module. If the visual system is a module then, it may turn out to be one that experience can wire into more than one region of the brain.

To summarize the state of play so far. Sterelny has identified three allegedly unique aspects of humans that, he claims, require flexible cognitive abilities for their origin and maintenance. Moreover, there appear to be considerable fitness benefits associated with cooperation, cumulative niche construction, and cognitive plasticity that would have outweighed the associated costs of developing the requisite neural hardware. However, as we have argued, these conclusions depend heavily on (a) idiosyncrasies in one's preferred interpretation of evolutionary events and (b) the nature of the task analysis associated with those events. Not only is it possible to come up with equally plausible evolutionary scenarios for how cooperation originated or how culture

evolved, but it is also often possible to provide a module-based account of the underlying architecture that supports those functions. So, how can we decide between Sterelny's Darwinian theory of psychology as learned and the evolutionary psychologists' theory of psychology as largely innate? Or are they both inevitably "just so stories" between which no evidence based choice is possible? Sterelny's evidence consists in some archaeology and primatology, a certain amount of contemporary biological anthropology and a great deal of independent theorizing by others about human evolution. We think that more direct sources of evidence might decide this issue. The two most promising areas of inquiry that might shift the evidential weight one way or the other seem to be neuroscience and genomics. In the remainder of this review we make some suggestions about what sort of evidence these disciplines could potentially offer.

Let's first consider what neuroscience suggests. It is a surprising discovery from the study of London cab-drivers that the hippocampus, the region in which spatial information is stored, actually enlarges under high demand (Maguire *et al.*, 1997). Similarly, the parts of violinists' brains responsible for fingering show enlargement compared to non-violinists (Elbert *et al.*, 1996). If other brain regions are like these regions, it may be the case that there are important psychological modules that are the products of *learning*, and not innate at all. There are other data from neuroscience which seem to support particulars of Sterelny's account. Recall for example, the role of hereditary perceptual priming as an alternative to domain specific innate modules. And now consider recent research on autism. fMRI studies reveal that patients with Asperger's syndrome (high function, normal intelligence autism) show lower activation in the medial prefrontal regions of the brain, where automatic processes take place, and higher activation in ventral regions—where inferential reasoning takes place—than normal individuals. This may suggest that normal individuals have an encapsulated theory of mind module which is defective in autistic individuals. However, it is also well established by neurological studies that the brain typically shifts problems from conscious processing regions to automatic processing over time, streamlining it, making it more encapsulated, non-conscious. In short, modules may be created in the brain as a result of experience. Now, on Sterelny's scenario, instead of natural selection for a theory of mind, there was natural selection on perceptual fine tuning for salient facial features. Interestingly, in persons suffering from hereditary autism, the defect is to be found in some gene coding for a perceptual capacity. On Sterelny's view, a theory of mind, or at least an ability to predict the behavior of conspecifics, is produced by scaffolded learning, and then shifted to automatic processing.

Furthermore, in high-function (Asperger's) autism, scaffolded learning enables the agent to substitute inferential reasoning to make at least some of the same predictions about behavior, as the fMRI evidence suggests. More important, however, current evidence suggests that the hereditary defect in autism is a break down in the perceptually primed eye-tracking of faces! The fact that high function victims of autism can compensate for this perceptual deficit by substituting conscious reasoning processes, provided scaffolded learning is available, would constitute neuroscientific evidence favorable to Sterelny's theory.

In spite of its limitations, research on neuroanatomy may eventually shed light on the dispute between the nativist modularity thesis and the empiricist central processor thesis. But there is another source of data besides neuroanatomy that could shed light, more immediately and with greater intensity on these competing hypotheses. This is the emerging area of neurogenomics. Surely, without grounding in genomics, no claim of innateness can really be established, still less can a claim about the adaptational etiology of a hard wired mental module be supported. And to explain how these modules make us different from our primate cousins, these genes must show characteristic marks of selection. By the same token, Sterelny's alternative scenario will also have implications for comparative genomics as well. What will these different implications be, and is there already evidence that vindicates one of these views more strongly than the other?

If there are innate mental modules, then must there be genes dedicated to their construction in development, a gene for folk-biology, or mate selection, or cheater detection, or grammaticality? The answer is almost certainly that there are no such genes for modules. If modules are like most other specialized structures of the body, they are the result of developmental programming by a large suite of genes many of which are also to be found in the suites of genes programming other structures as well. But if there are suites of genes, large and heterogeneous in number and developmental role, which do program⁶ the development of these modules, and their persistence is owing to their adaptational value, then these suites of genes must also be packaged together by selection in the more adapted lineages.

Of course to demand that nativists identify these genes and establish their co-evolution is vastly premature, given current information about the human genome. Nevertheless, there will be a difference in principle in the implications for currently available genomic data between nativist hypotheses and empiricist ones. If the differences between us and the chimps is a matter of distinct modules that most humans have and

⁶ This controversial claim needs defense. It is provided in Rosenberg 2006.

that most chimps lack, then there should be systematic differences between humans and chimps among the genes preferentially expressed in the brain. On the other hand if the difference between us and chimps is a matter of increased central processing power (plus the package of plasticity, group selection, niche construction), then we should expect another result when we compare gene sequences: a difference from other primates that results in the much increased size of the *Homo sapiens* cerebellum, but not in its functional differentiation.

The standard impression one gets from the current state of functional genomics is that we are a long way from understanding the various developmental roles of all but a handful of genes. This might be taken to imply that genomics has a long way to go before it can weigh in decisively on the nativist/empiricist debate. However, we think that this view is uninformed. Information currently available or available in the next few years may provide significant if not decisive evidence one way or the other. Findings in neurogenomics should be able to shed light on human evolution long before it has identified all, most, or even many of the genes responsible for neurological development and for the macromolecular chemistry that regulates brain processes. The very recent development of “gene-chips,” microarrays, will enable the neurogeneticist to identify the regions of the human genome in which the relevant genes for neural development and brain function are located.⁷ Once these regions are located and sequenced in whole or in part, comparisons with other regions of the human gene sequence, and with the gene sequences of other species, especially primates, can tell us something important about human cognitive evolution, some of it bearing on the nativist/empiricist dispute.

What has only recently been learned from sequence comparisons about the evolution of the human brain shows the promise of genomics for ultimately deciding this dispute. B.T. Lahn and co-workers have already taken the first steps in this research program (Dorus *et al.*, 2004). Of the many thousands of genes expressed in the brain, Lahn’s laboratory compiled a list of 214 already known to have a role in brain

⁷ The steps are relatively simple: extract the messenger (m)RNA from brain cells. Construct complementary (c)DNA strands from the brain mRNA. These cDNAs will preferentially bind to the gene sequences which coded for the brain cell’s mRNAs, arrange these cDNA sequences on a tile—a gene chip or microarray, wash cDNA from another organism of the same or different species over the chip. Those cDNA’s which preferentially bind to the sequences on the chip show that the same gene is expressed in the comparison organism as well. Those which fail to bind represent genes not expressed in the brains of both organisms. Thus we can quickly build up comparisons of large numbers of neural gene sequence differences among many different organism. Though we don’t know what these genes do in the brain cells, we know they play a role in neural development or brain-cell operation.

development or operation, either because their mRNA is expressed predominantly or exclusively in the brain, or because they are implicated in diseases of the brain. Of course exactly what their roles are in brain development and operation remains unknown. But they are known to be scattered randomly throughout the genome. Locating the orthologous sequences in humans, macaque monkeys, rats and mice, Lahn *et al.* compared the four sequences of each gene for synonymous nucleic substitutions and nonsynonymous nucleic acid substitutions.⁸ Synonymous substitutions will be evolutionarily invisible neutral mutations, while non-synonymous substitutions will be evolutionarily significant: they will change the role or the effectiveness of gene products. Such mutations are usually deleterious, but non-synonymous substitutions provide much of the variation that adaptation, and evolution require.

When Lahn *et al.* (2004) compared the ratio of nonsynonymous to synonymous substitutions in neurally expressed genes, they discovered that the proportion of adaptively significant nonsynonymous sequences was much greater in the primate species than in mice and rats, and proportionately even greater in *Homo sapiens* than in macaque monkeys (and as they later showed much greater in humans than in chimps). By comparison, the differences in ratio of synonymous to nonsynonymous sequence variation between primates and rodents in these genes was much greater than was the difference in the nonsynonymous/synonymous ratio for about 100 well-understood “house-keeping” genes that primates share with most other eukaryotic species. Lahn *et al.* conclude that there has been much stronger selection among the genes implicated in brain development and operation than among other genes controlling primate development and physiology, and even stronger selection among such genes in *Homo sapiens* than in macaques and chimps! Of course 214 genes is a very small proportion of the perhaps 20,000 genes predominantly expressed in the nervous system, but if this pattern of non-synonymous substitution differences is maintained, it will be increasingly implausible to suppose that human/intra—human psychological differences are a matter of one or a small number of gene-products present in humans and absent elsewhere.

If the gene-sequence differences between us and other primates is related largely to matters of cerebellum size, as opposed to structure or distinctive neural processes, innate modularity hypotheses will be hard-

⁸ Owing to the redundancy of the genetic code, the same gene-product can be produced by a number of different sequences of nucleic acids. Nucleotide differences in two distinct sequences that produce a gene product with exactly the same biological role are termed “synonymous”. Nucleotide differences that result in products with different biological effects are termed “nonsynonymous”.

er to support. To see why consider the next step in Lahn *et al.*'s study: Based on the admittedly limited information about their functional role, the 214 genes were categorized into ones "biased towards nervous system development"⁹—by loss of function in development following knock out, for example, as opposed to those "biased towards routine physiological operations"¹⁰—through pharmacological and biochemical evidence, and a third category of genes not classified either way because of insufficient functional data. The 53 genes implicated in development across the four species show much higher rates of nonsynonymous substitution than the 95 genes implicated in operation of the nervous system. (The 66 genes in the third category's rate of non-synonymous substitution was intermediate between the developmental and the operating genes). This is what one would expect if the differences between species, especially closely related species with little over-all sequence differences, were concentrated in the regulatory genes that control development. But it turns out that of the 24 gene sequences which showed the highest ratio of non-synonymous to synonymous substitutions, 25% are implicated in brain size differences among species.¹¹ In other words, nature has more strongly selected for brain size in recent evolution than for other differences between us and other primates or for that matter other mammals.

One of these most rapidly evolving genes in particular, *ASMP*, was already known to be particularly significant in its impact on brain size. To begin with, *ASMP* was first identified in a screen for genes associated in humans with microencephaly—reduced brain size. The gene turns out to be homologous to genes in creatures as evolutionarily distant from us as *Drosophila*, where it's malfunction reduces the amount of neural tissue in the fly. The gene's structure is highly repetitive, with multiple copies of a calmodulin binding region, whose number correlates with brain size from the worm (2 repeats) to *Homo sapiens* (74 repeats). In the human microencephaly is associated with a reduced number (68) of these regions. Finally, *ASMP* turns out to be even more rapidly evolving in humans since our last common ancestor [LCA] with the chimp! At least at present the evidence from molecular genetics suggests that the most highly selected evolutionary difference

⁹ Genes coding for developmental pattern signaling transcription factors, apoptosis, differentiation, migration, morphogenesis.

¹⁰ Genes coding for neurotransmitters, receptors, hormones, ion channels, synaptic vesicle components.

¹¹ A further 10 of the genes showing greatest proportion of non-synonymous substitutions are implicated, in the mouse at any rate, in defects in neural correlates of learning, circadian rhythm, conditioning, maternal and other unlearned social behaviors, as shown by knock-out studies.

between human brains and primate brains is brain size, and that brain size differences may be a relatively straightforward result of gene-sequence duplications. By contrast, there is little evidence of differences in the character or copy number of the genes that regulate processes throughout the brains of all the primates. If this pattern is born out as the number of genes increases by two orders of magnitude, then it will probably be safe to say that the difference between the human brain and other primate brains is not a matter of the evolution of distinct capacities programmed by distinct poly-genetic packages being selected together for their adaptational payoffs. Rather, the view of the human brain as differing by degrees from that of our primate cousins will be sustained, and the role of the environment in shaping this organ ontogenetically, instead of phylogenetically, will be strengthened. This is, of course, just what Sterelny would expect.

On the other hand, as we begin to amass more data about sequences implicated in brain development and operation, new evidence may strengthen modularity hypotheses. Once we have sequence data on a large number of genes for neural development and operation across several species, we will be able to employ the same technology for identifying synonymous and nonsynonymous nucleotide substitutions to date the appearance of these genes to periods before and after the LCA of humans and chimps 5 million years ago, before and after the LCA of humans and *Homo erectus* a million years ago, the LCA of humans and Neanderthals 600,000 years ago, and finally the LCA of humans and whatever lineage we separated from 200,000 years ago. Remember, to do this we won't need to know the actual functional role of the products of these genes, and we won't need to compare them to sequences from *Homo erectus*, Neanderthal, or other hominid species.¹²

Suppose we find that packages of neural gene sequences can be dated as emerging after our last common ancestor with, say, the Neanderthals, and suppose that studies of neurological diseases and deficits show a correlation between mutations in any one of the sequences in this package and a relatively discrete cognitive deficit, say the ability to interpret others' behavior, select mates, detect cheaters and free riders, identify dangerous botanicals, identify noxious stimuli after single trials, etc. Then we will have considerably strengthened the innate modularity hypothesis. This will be true even if the developmental roles of the genes in the package can be shown not to be regionalized, but spread to even distantly separated parts of the brain. Notice

¹² Though it would provide even stronger evidence to decide about the course of human psychological evolution if we can, and we may be able to, thanks to future developments in the amplification, sequencing and study of ancient DNA.

this is something else we can do long before we have identified the precise functions of the genes in such packages.

Whichever way neurogenomics plays out will have increasingly more impact on the conflict between nativist evolutionary psychologies and non-nativist ones. It is certainly too soon to decide matters between these competing theories, though there is surprisingly enough some genetic evidence for Sterelny's non-nativist alternative. What is most important to conclude is that in the relatively near future no one will be able to treat this dispute as an empirically irresolvable one between just-so-stories. And that makes Sterelny's up-dating of the empiricist's *Tabula Rasa* of very great importance.

References

- Bingham, L. 1999. "Human Uniqueness: A General Theory," *Quarterly Review of Biology*, 74: 133–69.
- Bingham, L. 2000. "Human Evolution and Human History: A Complete Theory," *Evolutionary Anthropology*, 9: 248–57.
- Brandon, R.N. 1990. *Adaptation and Environment*, Princeton, N.J.: Princeton University Press.
- Cosmides, L. and Tooby, J. 1994. "Origins of Domain Specificity: The Evolution of Functional Organization." In L. Hirshfeld and S. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge: Cambridge University Press.
- Dorus, S., Vallender, E., Evans, P., Anderson, J., Gilbert, S., Mahowald, M., Wyckoff, G., Malcolm, C. and Lahn, B.T., 2004. "Accelerated Evolution of Nervous System Genes in the Origin of *Homo Sapiens*," *Cell*, 119: 1027–1040.
- Edelman, G. 1987. *Neural Darwinism: The Theory of Neural Group Selection*, New York: Basic Books.
- Elbert, T., Sterr, A., Candia, C., Rockstroh, B. Taub, E. 1996, "Representational Cortical Plasticity as Revealed by Magnetic Source Imaging: How the Brain Learns to Play the Violin," *First Berlin Workshop on Cortical Plasticity*, Berlin, November.
- Fletcher, G.J. and Stenswick, M. 2004. "The Intimate Relationship Mind," In K. Sterelny and J. Fitness (eds.), *From Mating to Mentality: Evaluating Evolutionary Psychology*, London: Psychological Press, pp. 71–93.
- Fodor, J.A. 1983. *The Modularity of Mind: An Essay in faculty Psychology*, Cambridge: MIT Press.
- Frank, R.H. 1988. *Passions Within Reason: The Strategic Role of the Emotions*, New York: Norton & Co.

- Gangestad, S. and Simpson, J. 2000. "The Evolution of Human Mating: Trade-Offs and Strategic Pluralism," *Behavioral and Brain Sciences*, 20: 573–644.
- Godfrey-Smith, P. 1996. *Complexity and the Function of Mind in Nature*, Cambridge: Cambridge University Press.
- Gray, R., Heaney, M. and Fairhall, S. 2004. "Evolutionary Psychology and the Challenge of Adaptive Explanation." In K. Sterelny and J. Fitness (eds.), *From Mating to Mentality: Evaluating Evolutionary Psychology*, London: Psychological Press, pp. 247–268.
- Horan, R.D., Bulte, E. and Shogren, J.F. 2005. "How Trade Saved Humanity from Biological Exclusion: An Economic Theory of Neanderthal Extinction," *Journal of Economic Behavior & Organization*, 58: 1–29.
- Maguire, E.A., Frackowiak, R.S.J. and Frith, C.D. 1997. "Recalling Routes Around London: Activation of the Right Hippocampus in Taxi Drivers," *Journal of Neuroscience*, 17(18): 7103–7110.
- Plotkin, H. 1994. *Darwin Machines and the Nature of Knowledge*, Cambridge, Mass.: Harvard University Press.
- Rosenberg, A. 2006. *Darwinian Reductionism or How to Stop Worrying and Love Molecular Biology*, Chicago: University of Chicago Press.
- Rozin, P. and Fallon, A.F. 1987. "A Perspective on Disgust," *Psychological Review*, 94: 23–41.
- Samuels, R. 2000. "Massively Modular Minds: Evolutionary Psychology and Cognitive Architecture." In P. Carruthers and A. Chamberlain (eds.), *Evolution and the Human Mind*, Cambridge: Cambridge University Press.
- Schmitt, D.P. 2005. "Sociosexuality from Argentina to Zimbabwe: A 48 Nation Study of Sex, Culture and Strategies of Human Mating," *Behavioral and Brain Sciences*, 28: 247–311.
- Simpson, J. and Orina, M., 2004. "Strategic Pluralism and Context Specific Mate Preferences in Humans." In K. Sterelny and J. Fitness (eds.), *From Mating to Mentality: Evaluating Evolutionary Psychology*, London: Psychological Press.
- Sober, E. and Wilson, D.S. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, Mass.: Harvard University Press.
- von Melchner, L., Pallas, S.L. and Sur, M. 2000. "Visual Behaviour Mediated by Retinal Projections Directed to the Auditory Pathway," *Nature*, 404: 871–876.