Three Evolutionary Precursors to Morality

JOSEPH HEATH *University of Toronto*

ABSTRACT: Recent evolutionary thinking suggests that morality may have three distinct roots, corresponding to the three "master concepts" in the history of philosophical ethics: sympathy, virtue, and duty. Evolutionary biologists are nearly unanimous in the view that there are two distinct mechanisms that can generate altruism in other animal species, and there is increasing recognition that there must be a third mechanism at work among humans. Altruism in the biological sense is not the same thing as morality, but it is a precursor to morality. An examination of such precursors can help us gain a deeper understanding of morality proper.

RÉSUMÉ: Les travaux récents sur l'évolution suggèrent que la morale pourrait avoir trois racines distinctes, correspondant aux trois «concepts maîtres» dans l'histoire de la philosophie éthique : la sympathie, la vertu et le devoir. Les biologistes évolutionnistes s'accordent presque unanimement pour dire qu'il existe deux mécanismes distincts qui peuvent générer l'altruisme chez les autres espèces animales et l'on reconnaît de plus en plus qu'il doit y avoir un troisième mécanisme en jeu chez les humains. L'altruisme au sens biologique n'est pas la même chose que la morale, mais il en est un précurseur. Une étude de tels précurseurs peut nous aider à mieux comprendre la morale en tant que telle.

One of the unspoken assumptions quite widely shared among moral philosophers is the belief that human beings have a unified moral pyschology. Roughly speaking, morality involves action that is, at least prima facie, contrary to self-interest. This generates two immediate problems. The first involves determining whether moral action, under this description, is possible, and if it is,

***Dialogue* 48 (2009), 717-752.**
© Canadian Philosophical Association/Association canadienne de philosophie, 2009
doi:10.1017/S0012217309990400

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

explaining how such action might come about (e.g., what the underlying intentional structure might be). The second involves the normative task of justifying a moral course of action to an agent who, perhaps disposed to act this way, nevertheless wonders why he should not revise his goals to assign greater priority to self-interest. Attempts to address the second problem typically — although not always — presuppose some answer to the first problem, since it is difficult to know how to address an argument to a person without having at least a general sense of which intentional states of that person one is seeking to modify. Thus no matter how "metaethical" the reflections of moral philosophers may become, their theories are typically structured by a (more or less explicit) picture of how moral reflection gets translated into action. In other words, they all presuppose some characterization of the psychological mechanism that enables us to expand our circle of concern beyond self-interest, narrowly construed. Unfortunately, it is often the normative dimension — the desire to justify moral obligations — that drives the project, leading philosophers simply to posit the psychology that seems to them most conducive to the construction of such an argument.[1] And since the normative argument usually has a fairly linear structure, it is necessary to posit only one psychological mechanism; as a result, only one gets posited.

There are three "moral psychologies," constructed more-or-less along these lines, that have dominated the history of philosophical reflection on the subject. The first is associated with the tradition known as moral sentiment theory, and can be thought of as an approach that privileges the concept of *sympathy*. The central phenomenon here is the way that the suffering of one tends to generate distress among others. From this, it is not so difficult to imagine that we act morally because we *care* about others (and that "sympathy," or some related concept, identifies the mechanism through which the concerns of others are communicated). The second approach, usually associated with Aristotle, focuses on the concept of *virtue*, understood as a disposition — either cultivated or chosen through practical deliberation — that expands the scope of the individual's goals in a way that encompasses certain concerns of others. Finally, there is the concept of *duty*, which has occupied a central role in philosophical discussion since Kant. The central phenomenon here is that of *rule-following*, viz. the fact that individuals often perform actions because of some characteristic of the action itself, independent of its consequences. Thus individuals may act contrary to self-interest not because they have adopted some other goal, but simply because they regard the *action* as obligatory.

Any effort to combine the insights of these different traditions — to declare, in effect, that they each articulate some aspect of the phenomenon — usually comes across as little more than unprincipled eclecticism. Yet the dogged persistence of these different perspectives, combined with their rather self-evident plausibility with respect to their own paradigm cases, suggests that it may be time to take more seriously the idea that "morality" is not a unified phenomenon, that each of these "moral psychologies" may correspond to a distinct psychological

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

mechanism capable of motivating individuals to transcend their self-interest in a different way.[2] In this paper, I would like to argue that an evolutionary perspective renders highly plausible the suggestion that morality may have three distinct psychological roots (even for those who reject "sociobiological" approaches to the study of culture and cognition). Evolutionary biologists are nearly unanimous in the view that there are two quite distinct mechanisms that can generate "altruism" among other animal species — and there is increasing recognition that there must be a third mechanism at work among humans. Thus insofar as our psychology is a product of evolutionary adaptation, there is good reason to expect that it will bear the imprint of these three distinct mechanisms. Of course, altruism in the biological sense is not the same thing as morality, but it is, I would like to suggest, an important *precursor* to morality. Furthermore, an examination of these evolutionary precursors can serve as a useful heuristic for gaining a deeper understanding of morality proper.

Mental Adaptation

I would like to begin by sketching out the broad contours of the philosophy of mind that I happen to believe is correct. It is known most commonly as the "dual-process" view.[3] I say "happen to believe" because I am not in a position to defend this view here; I merely want to state it, in order to clarify the framework within which the subsequent reflections on the subject of altruism and morality will be developed. (I will of course provide some mention of the arguments that I think support this view, but merely as a way of rendering it intelligible.)

One of the most significant developments in philosophy of mind in the past three decades has been increased recognition of the importance of *modularity* in explaining certain aspects of human cognition. While the term is controversial in some quarters (in part due to overuse by more zealous proponents), the central phenomenon is widely accepted with respect to "peripheral" (as opposed to "central") systems. Consider an ability like facial recognition. The fact that we are able to recognize people again, after having seen them only once, perhaps fleetingly, is remarkable. The fact that we are able to recognize people again, ten or twenty years after having last seen them, is positively astonishing since, strictly speaking, they look completely different (their noses and ears will have grown, their faces broadened, their teeth changed, etc.) Furthermore, we have no idea how we do this (and have had only modest success programming computers to do it); people just "look familiar." Thus facial recognition has all of the hallmarks of a cognitive module: the processing is fast, functionally specialized, requires little or no attention, and uses a set of processes that are unavailable to consciousness.[4]

What is noteworthy about such modules is that they seem to be fairly clear candidates for adaptive evolutionary explanation. On the one hand, modularity seems to be precisely the sort of structure that would be favoured by evolutionary processes, which must rely upon incremental changes that are nevertheless capable of delivering improvements in fitness.[5] Highly specialized

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

heuristics seem like the sort of thing more likely to evolve than completely general problem-solving systems. On the other hand, the type of cognitive abilities we have that exhibit the most clear-cut signs of modularity seem to be ones that would have offered obvious benefits in the environment of evolutionary adaptation (e.g., anticipating trajectories, determining when food is spoiled, detecting anger or fear in others, etc.) The two factors combine to lend considerable plausibility to the hypothesis that an important range of cognitive processes involve modularity in this sense.[6]

Layered over this set of cognitive modules (sometimes known as the "adaptive unconscious"[7] or "the autonomous set of systems [TASS]"),[8] we have the "analytic" or the intentional planning system. This is the seat of conscious thought, or what Robert Brandom refers to as "sapience" (as opposed to "sentience").[9] The most important characteristic of this type of thought is that it is linguistically structured. The fact that it is language-dependent means that the type of processing done at this level has a very different structure from that which occurs at the level of the adaptive unconscious. (As Daniel Dennett puts it, language permits the implementation of a "serial virtual machine" "on the parallel hardware that evolution has provided for us").[10] Thus the intentional planning system is *not* a cognitive module (or is not usefully described as a module). The processing done by this system is linear, non-task-specific, highly flexible, and subject to self-correction.[11] But because of the nature of the underlying hardware, the implementation of this system is also "inefficient."[12] As a result, intentional reasoning is slow, very demanding in terms of cognitive resources (especially attention), and has available limited "on-board" memory. This is why computers — designed "from the ground up" for serial processing — are able to accomplish with only trivial computational resources certain reasoning tasks that an unaided person finds very difficult. This also suggests that the intentional system is not "adapted" in the same sense that other cognitive modules are, but is a product of an adaptation that essentially colonized existing cognitive resources and redirected them to some other use. It is an "exaptation," or as Dennett puts it, "a *very* recent and rushed add-on, no doubt an exploitation of earlier sequencing circuitry."[13]

The most plausible hypothesis, in my view, is that language began as an external communication device, but that speech acts developed a compositional structure that allowed them to serve also as a computational system.[14] Thought, in the sense of sapience, began as the capacity to mentally rehearse a sequence of speech acts before performing them.[15] Through a process of increased internalization, speech eventually became the "vehicle" for a style of cognitive processing that is quite distinctive, relative to the native architecture of the brain.[16] To take just one example, it appears that human infants share with other primates two basic heuristics for handling numerosity: the first a "subitizing" system for making judgments involving a very small numbers of objects, and the second a "large number" system for making approximate judgments concerning large collections of objects. Perhaps surprisingly, human infants do not perform

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

much better than do other primates when it comes to these heuristic judgments.[17] Yet using linguistic resources, humans acquire the ability (at about age four) to count not just to 50 or 100, but as high as they like. This is a qualitative, not merely a quantitative, improvement in performance, one that appears to be based upon a unique capacity to perform recursive operations.

It is observations such as these that fuel the speculation that language is not merely "one more module" being added to the heap, but that its introduction represents a more fundamental transformation of our prior cognitive capacities.[18] From this perspective, it is only states of the intentional planning system that are linguistically structured, or that possess propositional content. This makes it a category error to assign semantic or intentional content to states of the adaptive unconscious. When we talk about an agent having "beliefs" or "desires," and we take these to designate intentional states, then these refer only to states of the intentional planning system. Naturally, there are computational states of the adaptive unconscious that serve similar functions with respect to particular modules, but there is a good case to be made for using some other set of terms to describe them. Prelinguistic infants and other primates are able to notice the difference between two objects and three objects and to treat these sets differently, but there is an important sense in which they lack the *concept* of number, because they lack mastery of the counting procedure. Similarly, we may be able to observe a three-dimensional object and form a mental picture of it, and it may even make sense to talk about its being a representation of something in the world. But it invites terrible confusion to refer to it as a "belief," or to try to understand beliefs on the model of mental pictures. The same can be said for desires. A person who ducks when she sees an object entering her peripheral field of vision can be described as wanting to get out of the way. But the type of goal-state involved in this sort of autonomous response is very different from a full-blown "desire." In particular, it seems essential to the concept of a desire that it stand in inferential relations, or relations of logical dependence, to a host of other desires (which an impulse to duck manifestly does not).

One of the most important distinctions between cognitive operations performed at the level of the adaptive unconscious and those performed by the intentional planning system is that various mechanisms associated with the former seem to have only limited capacity for self-correction, or modification of their routines. Daniel Kahneman and Amos Tversky, for instance, have shown that much of our everyday probabilistic reasoning takes as its point of departure heuristics contributed by the adaptive unconscious.[19] When asked to estimate the likelihood of an event, for instance, our first response seems to rely upon an "availability" heuristic, whereby the ease with which a similar event can be recalled from memory is used to estimate its probability (or frequency).[20] This leads to a number of well-known cognitive biases, such as a tendency to overestimate the probability of events that have left a more powerful impression upon us or that we have heard about more often. Yet it is noteworthy that the way we *correct* these tendencies is through explicit symbolic representation of

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

the problem. We are unable to change the way that our "first guess" at a solution is generated (i.e., the underlying heuristic), yet in the case of probability estimates, we also generally are unable to keep an explicit record of the multifarious events that occur in our immediate environment. Thus we develop a conscious work-around to the "bugs" that exist in our hardware by forming an explicit representation of the circumstances that are likely to cause this heuristic to misfire, thereby imposing a filter upon our own judgments, either modifying or suspending confidence in those that are likely to be biased. So when asked to guess which is higher, the murder rate or the suicide rate, or which is more dangerous to children, a gun in the home or a swimming pool, anyone familiar with her own cognitive biases will immediately be suspicious of the inclination to pick the first of either pair.

It is precisely the ability of the intentional planning system to override the adaptive unconscious and to develop work-arounds to the limitations of its heuristics that has allowed human beings to survive and flourish outside the environment of evolutionary adaptation.[21] Many evolutionary psychologists, unfortunately, have an almost perverse inclination to regard anything genetic as good. As a result, "they too easily dismiss the role of general intelligence and/or general computational power in overriding deleterious TASS responses."[22] This is an important point for the reflections that follow, because I intend to argue that the three precursors to morality represent only the output of evolutionarily adapted cognitive mechanisms. They provide guidance to our actions in very much the same way that the innate heuristics studied by Kahneman and Tversky provide guidance to our judgments. This is why they are only precursors to morality. They provide us with inclinations, behavioural or reactive dispositions, "gut feelings," prime facie judgments, and cognitive schemata. But what we choose to *do* with all this, at the level of the intentional planning system, is a separate question. The outputs of the adaptive unconscious *influence* our desires in myriad ways, but they are not themselves desires. As a result, I will not be providing an evolutionary or adaptationist account of our desires (in the style of David Sloan Wilson and Elliot Sober).[23] On the contrary, if the theory of mind being presupposed here is anywhere near correct, then this style of evolutionary explanation involves a serious conceptual error.

Alturism in Nature

The best place to start, when developing an evolutionary approach to morality, is with the puzzle of altruism. "Morality," which we may stipulatively define as intentional action that is contrary to the interests of the actor but advances the interests of others, is a phenomenon that is specific to humans (insofar as only humans have a set of intentional goals).[24] Altruism is a more general term, used in evolutionary biology to refer to behaviour that decreases the (reproductive) fitness of the individual who performs it, but increases the fitness of one or more others. Unlike morality, many different species exhibit altruism. Among humans, there is substantial overlap between morality and altruism but there

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

are also important divergences, since many of the goals that we choose to pursue reduce our reproductive fitness. Of course, how it is that we acquire the capacity to pursue goals that reduce our own fitness is also something of a puzzle — a puzzle not unrelated, it turns out, to the one concerning altruism. Thus it is best, when attempting to piece together this story, to begin at the beginning.

It is only very recently that altruism has come to be seen as an evolutionary puzzle. In the early half of the twentieth century, there was a widespread tendency among evolutionary biologists to regard particular traits as adaptive merely because they were "good for the species," even though it was widely understood that individual-level selection would tend to work against the development of such traits.[25] For example, an individual (gopher, vervet monkey, etc.) who produced an alarm call upon spotting a predator would also be likely to attract the attention of this predator, thus reducing his own fitness relative to that of others.[26] Nevertheless, it was felt that because groups that contained individuals exhibiting altruistic traits would do so much better than groups that lacked such individuals, natural selection would still tend to favour the development of altruism. This conviction rested in part upon an unwillingness to countenance the possibility of massive suboptimality in the outcome of evolutionary systems.

These assurances were disrupted in 1966 by George Williams, who argued that group-level selection effects would almost always be weaker than those occurring at the individual level.[27] While groups that contained altruists might do better than groups that lacked altruists, a disproportionate share of these benefits would flow not to the altruists within the group, but to the non-altruists. Thus what group-level selection tends to favour is not "being an altruist," but rather "being in the close company of an altruist." This has the structure of an evolutionary prisoner's dilemma, leading individually advantageous "free rider" traits to be systematically favoured, even when they are extremely damaging to the species as a whole. Thus altruism must be regarded as very much the exception, not the baseline.

To see a clear-cut example of this, consider the sex ratio in a typical mammalian species. From the standpoint of fecundity, what matters is the size of the female cohort. Eggs are costly, whereas sperm is cheap (by definition, as it turns out). So when it comes to reproductive fitness, the sex ratio that is best from the standpoint of the *species* would be very heavily weighted toward females (perhaps 99:1 female to male). By contrast, a 1:1 ratio is extremely wasteful from the standpoint of the species. It generates a huge population of "surplus" males, most of whom are not needed — and in many species, most of whom are not even used — for the reproduction of the species. At the same time, these males consume resources that would otherwise be available for members of the population involved in reproduction, fight with other males for reproductive opportunities, and so on. Thus the 1:1 sex ratio is suboptimal, in the sense that it produces a much lower level of average fitness for members of the species than would a ratio more skewed toward females. How could this be?

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

The answer lies in the fact that natural selection does not promote average fitness, it only promotes individual fitness.[28] A 99:1 sex ratio is not evolutionarily stable. While it may promote a high degree of average fitness, it is also susceptible to deviation. The first male born with a tendency to produce two sperm containing Y chromosomes, rather than just one, out of every 100, essentially doubles the number of grandchildren he can expect to have (and thus, the number of offspring likely to possess this trait). Yet this is a race to the bottom. As more individuals acquire this trait, fecundity drops and average fitness declines. Thus these individuals generate more male offspring but at the same time, confer upon each one a diminished probability of successfully reproducing. This continues until the number of males being born outnumbers females, in which case producing more female offspring becomes individually advantageous. So things settle out at around 1:1.

The example of sex ratio is useful because it is massively suboptimal, yet ubiquitous. This shows, in turn, the power of the evolutionary forces that are lined up against the development of altruistic traits. Furthermore, it shows that human beings are not exempt from these general strictures at any deep biological level. If we were, there would be a lot more women around. Thus altruism needs to be understood as an exception, one that emerges against a background that systematically favours self-interested, or "anti-social," behaviour.

There are, however, two mechanisms that are widely regarded as having the ability to generate such exceptions, along with a third that is currently the subject of widespread discussion:

1. Inclusive fitness

The first and most well-known mechanism is what W. D. Hamilton refers to as "inclusive fitness," or "kin selection."[29] The existence of this mechanism turns out to be, perhaps ironically, a fairly direct consequence of so-called "selfish gene" theory.[30] Not only do the "interests" of the individual not always align with the "interests" of the group, but the "interests" of the gene do not always align with the "interests" of the individual. Every gene reproduces in an environment that contains multiple copies of that same gene, housed in other individuals. Genes are fundamentally indifferent between reproducing themselves and reproducing the copies of themselves found in others. Thus a trait that reduces the fitness of one individual, but provides a greater increase in fitness for some other individual who possesses that trait (or provides a lesser increase to a greater number of such individuals), will be favoured by natural selection.

The clearest example of this is parental investment in offspring. Imagine a species of bird that has no tendency to feed or care for its young, but simply lays eggs and leaves. If a mutation occurs that makes the parent more inclined to stick around and feed its young, this mutation may spread quite rapidly in the population. Even though the presence of this gene will decrease the fitness of the parent (by increasing its energy expenditure while decreasing its food intake), there is a fifty per cent chance that each of the offspring has that gene,

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

and since each will have its fitness dramatically improved through the parent's efforts, the gene will prosper. In a sense, the gene sacrifices the copy of itself found in the parent, in order to promote the multiple copies of itself found in the offspring. When the notion of "individual fitness" is expanded in order to take these indirect effects into account, it yields the notion of "inclusive fitness."

In this respect, Richard Dawkins's *The Selfish Gene* is rife with rhetorically misleading passages. The title is often misread as implying that there is a "gene for selfishness." What it really means is that natural selection promotes selfishness only at the level of the gene, *not* at the level of the individual (or only indirectly at the level of the individual). The fact that the "interests" of a gene will typically be promoted by the individual pursuing his own interests is merely an empirical generalization, not a claim about the inner dynamic of evolutionary systems. Dawkins does state this in various ways, yet it still takes considerable mental effort to read sentences such as the following without misunderstanding: "Like successful Chicago gangsters, our genes have survived, in some cases for millions of years, in a highly competitive world. This entitles us to expect certain qualities in our genes. I shall argue that a predominant quality to be expected in a successful gene is ruthless selfishness. This gene selfishness will usually give rise to selfishness in individual behaviour."[31] The word "usually" in the final sentence is of the utmost importance. It is crucial to remember that the tenderness and compassion with which a mother cares for her young counts as an example of "ruthless selfishness" at the level of the gene (where the mother is essentially cast aside as a used husk, in order to promote the fitness of her parasitic brood). Thus the way Dawkins uses the word "selfish" is rather technical, and bears only a tenuous connection to the ordinary English sense of the term.

Inclusive fitness can also explain why siblings and relatives in many species often help one another. While the chance that one shares a particular gene with one's child is fifty per cent, the chance of sharing it with one's brother or sister is also fifty per cent, and thus the chances of sharing it with one's nephew or niece is twenty-five per cent. Because of this, saving four of one's brother's children is equivalent, from an inclusive fitness point of view, with saving two of one's own, which is equivalent to saving oneself. Of course, the sort of altruism that can be sustained in this way is limited. In the best-case scenario (thanks to sexual reproduction) the benefit at the receiving end of the altruistic act must be at least twice as great, in terms of fitness, as the cost to the individual performing it. Furthermore, if individuals have no way of discriminating between kin and non-kin, the pattern may not be sustainable. (Thus magpies, in order to defend themselves against the great spotted cuckoo, will destroy any egg that they do not recognize as their own. Since their ability to distinguish one from the other is limited, this requirement results in a reduction in fitness — but evidently less so than being suckered by the cuckoo.)[32]

It is important to remember that relatedness is not actually the central phenomenon with this form of altruism (which is why the term "inclusive fitness" is in certain respects preferable to "kin selection"). What permits the evolution of

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

altruism is simply *correlation of strategies* (to use the language of evolutionary game theory).[33] An altruistic trait may be evolutionarily successful if there is some characteristic of the trait that makes those who benefit from it more likely to be those who possess that same trait (and less likely to be those who do not). If "birds of a feather flock together," then that is sufficient to generate altruism. In this respect, kinship is simply one mechanism that can generate correlation. Under certain conditions group selection constitutes another.

Sexual reproduction, from this perspective, represents the primary obstacle to the development of altruism — because it reduces the probability of a particular gene's being shared between the most closely related individuals to only fifty per cent. A society of clones would be one in which natural selection favoured unlimited altruism, and unlimited individual sacrifice for others. For confirmation, one need only look at the level of co-operation exhibited among cells in a body (since a multi-cellular organism is essentially a society of clones). Every somatic (or non-germline) cell in our body is essentially sacrificing itself for the sake of the gametes. If one looks at the division of labor within the body, e.g., between liver cells and neurons, and compares it to the level of differentiation at the social level within a typical mammalian species, it provides some indication of the formidable obstacles that sexual reproduction creates for the evolution of altruism. From this perspective, what inclusive fitness represents is simply a partial attenuation of these obstacles.

2. Reciprocal altruism

The second primary mechanism known to promote altruistic behaviour is referred to, following Robert Trivers, as reciprocal altruism.[34] This form of altruism has a "you scratch my back, I'll scratch yours" structure. If the altruist can somehow provoke the beneficiary of his actions to return the gesture or to produce some other sort of benefit, then the net effect will be an improvement in fitness. Reciprocal altruism differs from inclusive fitness in that, with the latter, it is the copy of the individual's gene found in the beneficiary that benefits, whereas with the former, it is the individual's own copy that benefits, via the actions of the beneficiary of the initial act. The beneficiary, however, need not possess the same gene. This is most obvious in the case of altruistic interspecies symbioses (as with gobies, a species of "cleaner fish" who remove ectoparasites from the gills of grouper while the grouper, in turn, refrain from eating them). When altruistic behaviour benefits individuals belonging to another species, it is impossible for the gene to be benefiting some other copy of itself. It must be that the benefits of the behaviour somehow redound to the individual who performed it (and thus that the copy of the gene underlying the behaviour improves its own chances of reproducing).

Thus systems of reciprocal altruism may be adaptive insofar as they enable individuals to *co-operate* in fitness-enhancing ways. Yet this connection between altruism and co-operation has the potential to generate some confusion. After all, with inclusive fitness the individual genuinely sacrifices himself in

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

order to benefit another. With reciprocal altruism, on the other hand, the individual who performs the action ultimately benefits from it (making the strategy seem more like "clever selfishness" than genuine altruism). What makes it count as altruistic is the availability of a free-rider strategy that the individual forgoes. The grouper could wait until the gobie was finished removing ectoparasites, then eat it, yet he refrains from doing so. Thus co-operation involves a *failure* to enhance one's own fitness at the expense of another, which is what makes it altruistic.

Reciprocity is the form of altruism underlying the well-known "tit-for-tat" strategy in evolutionary game theory.[35] Tit-for-tat essentially instructs a player confronting a repeated prisoner's dilemma game with a single opponent to act co-operatively on the first round and in each subsequent round to do whatever the opponent did in the previous round (co-operate if the other co-operates, defect if the other defects). Thus two tit-for-tat players who encounter one another will be able to sustain co-operation, which in turn will confer significant benefits. Free riders, on the other hand, will get the "one shot" benefit of defection but will have to pay for it by having others defect upon them in future rounds. Thus they get limited benefits from free riding, and are denied access to the fruits of co-operation.

It is widely believed that this sort of conditional reciprocity underlies a range of altruistic behaviours among animals, such as food sharing, mutual grooming, and some alarm calls. Non-human primates, for instance, spend significant amounts of time (up to twenty per cent of the day) and energy grooming themselves and others. Although much of this occurs amongst kin, unrelated individuals also establish grooming relationships which they often maintain over long periods of time. Since primates are quite capable of distinguishing kin from non-kin, it is widely thought that reciprocal altruism, and not kin selection, is responsible for this behaviour. An individual may engage in "experimental" grooming of an unrelated individual. Reciprocation of the gesture increases the propensity to return to that individual and groom him or her again.

It should be noted that an altruistic disposition based on reciprocity, in order to be sustainable, must include not only a disposition to assist those who have assisted one in the past, but also a disposition *not* to assist those who have failed to assist one in the past. Altruists not only must want to be associated with other altruists, they also must want to dissociate from selfish agents. Again, this just means that there must be some correlation of strategies in the way that interactions occur, so that altruists are more likely to interact with other altruists and selfish agents more likely to interact with other selfish agents. An unconditional altruistic strategy is not evolutionarily robust even if there is some mechanism that increases the chances that it will be reciprocated, because the presence of such unconditional co-operators in the population enhances the fitness of the free riders more than it does that of the other altruists. Unless there is some mechanism that deprives the free riders of these benefits, they will steadily increase their share of the population at the expense of altruists.

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

(This is the primary reason that group selection promotes altruism only in cases where some of the groups contain *only* altruists. If all the groups are mixed, then the free riders always do best.)[36]

Incidentally, there is no reason that reciprocal altruism need involve the *same* individuals interacting with one another over time. Since correlation of strategies is all that is required, one can still sustain altruism in a repeated game in which partners are shuffled at the end of each round, so long as the "tit-for-tat" players do whatever the opponent did in the previous round *regardless* of who that opponent was playing against. If adequate information is communicated at the end of each round about what each player has chosen, then individuals can (at least in principle) develop reputations for being co-operators, and thus elicit greater co-operation from others. Following Richard Alexander, this was given the special name of "indirect reciprocity."[37] It is important to keep in mind, however, that indirect reciprocity does not involve any fundamental change in the mechanism that sustains co-operation.

Finally, it should be noted that there has been a recent surge of skepticism about the extent of genuine reciprocity in Nature. In part this is because reciprocity is not a very robust mechanism, so in many cases it is difficult to see how a system of reciprocity could ever get started in a population in which it is absent. A more common phenomenon seems to be so-called "by-product mutualism," in which traits arise that are individually advantageous yet also have spill-over effects for others. A closely related phenomenon is "pseudoreciprocity," in which an individual generates benefits for some other in order to enhance a by-product benefit.[38] An example would be butterfly larvae who feed ants, while the ants protect the larvae from predators. The behaviour of the ants, however, is one that is exhibited with *any* food source, and would be unchanged even if the larvae were to disappear. Assuming that the larvae have developed an adaptation that benefits the ants, this is still only pseudoreciprocity because the ants are not behaving altruistically (or more specifically, their strategy is not *conditional* upon reciprocation — even in the long-run evolutionary sense).

3. Cultural transmission

If one thinks of phenotype in the broad sense, to include a set of behaviours, then there are many species that exhibit what Robert Boyd and Peter Richerson refer to as "population-level" effects, where phenotype is transmitted directly from individual to individual or generation to generation without any modification of the underlying genotype. This is one way of describing cultural transmission. For example, there are several celebrated examples of "traditions" developing among songbirds, where unrelated individuals brought into the group will pick up on a pattern of song. This is not surprising, since cultural transmission is based upon a number of very obvious innovations, that can be found throughout the natural world. *Cumulative* cultural transmission, on the other hand, is comparatively rare, having only been convincingly demonstrated in the case of humans and songbirds (and with respect only to songs, in the case of the latter).[39]

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

The first step in establishing cultural transmission is simply developmental plasticity. In cases where an organism's environment is stable, natural selection will tend to favour a developmental trajectory that is "canalized" and thus invariably results in the trait that is the best (local) adaptation to that environment. In cases where the environment varies between state *x* and *y*, selection will favour a developmental trajectory that is "plastic" and thus results in some trait *a* under conditions *x*, but trait *b* under conditions *y* (where each trait represents the best adaptation to those conditions). The second major step is learning, which in its most primitive form is simply a feedback relationship between the environment and the developmental process such that traits or behaviour that are successful will be reinforced, while those that are unsuccessful will be extinguished. A fairly obvious extension of this is trial-and-error learning, in which the individual engages in some variation of the behaviour in order to discover the best form. Finally, there is social learning, where the individual looks to the behaviour of conspecifics for clues as to which variant will be best. This is also a fairly obvious adaptation, since it allows individuals to economize on learning costs. It is at this point that population-level transmission of phenotype becomes possible.

It is important to recognize, however, that social learning is essentially a free-rider strategy — someone else pays the price associated with failed attempts in trial-and-error learning, while the individual who copies her derives all the benefits associated with any success that she has obtained. When too many do this, the result is suboptimal. As a result, the number of social learners in a population can be expected to grow until their average fitness converges with that of trial-and-error learners, but with even a moderate level of environmental variability it is unlikely ever to exceed it.[40]

Furthermore, one cannot appeal to the benefits of cumulative cultural transmission (ranging from stone weapons to modern medicine) as an explanation for the adaptive value of whatever social learning system produces it, since the learning system had to have become prevalent in the population long before the emergence of such benefits. Thus the only way that culture-dependence can prosper is if the capacity for social learning also provides some up-front improvement in the efficacy of individual learning strategies (for example, if individuals establish a "prior" through observation of conspecifics, then selectively update through trial-and-error).

Obviously the emergence of these more complex learning rules significantly alters the evolutionary dynamics that govern reproduction of phenotype. In particular, one can see that it might allow all sorts of genetically maladaptive behaviours to emerge. For example, phenotypes that reduce reproductive fitness but are very good at attracting imitators may enjoy considerable success in the domain of cultural reproduction. As long as the overall benefits associated with the social learning system outweigh the cost imposed by the occasional acquisition of (genetically) maladaptive behaviour, the system as a whole will be favoured by natural selection. Because of this, it has been suggested that

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

cultural inheritance systems may provide a more propitious environment for the development of altruism than do genetic ones. Altruism, in other words, might emerge and be reproduced at the population level or as a culturally transmitted pattern of behaviour, in a way that bypasses the "selfish" genetic substratum. (Speaking loosely, the "selfish meme" may be less likely to promote individual selfishness than the selfish gene is.)

It is extremely important to realize, however, that the mere fact that a cultural inheritance system of this type arises does not in itself imply that there will be any divergence between culture and biology (and thus does not mean that there will be any room for the cultural transmission of genetically maladaptive behaviour). It all depends on how cultural traits are inherited. For example, if children engage in imitative learning only from caregivers, and for inclusive fitness reasons only parents and relatives act as caregivers, then culture will necessarily evolve in lockstep with biology. A particular pattern of learned behaviour will be passed along only if it increases the individual's chances of becoming a "role model" or "cultural parent" to a child. And since only biological parents become cultural parents, the spread of a particular form of learned behaviour will be determined entirely by its contribution to the reproductive fitness of the individual who learns it. Under such circumstances, the difference between culture and biology would be nugatory (and thus, cultural transmission would add nothing new to the story on altruism).

Somewhat more surprisingly, if infants select a cultural parent at random (i.e., through an unbiased sampling procedure) from the population, then there will also be no difference between biological and cultural evolution. This point is often overlooked by critics of sociobiology. With unbiased transmission, a learned behaviour pattern will be able to propagate only if it increases its representation in the population, which means increasing the longevity or fecundity of any individual who adopts it. While cultural evolution will be *faster* than biological evolution, there will be no culturally-sustained profile of behaviour in a population that could not *also* be sustained as the equilibrium of a purely genetic system of inheritance. By implication, all of the constraints that natural selection in the biological domain imposes upon altruistic behaviour would apply with equal force in the cultural domain.

These constraints are important to keep in mind, because many philosophers (and subpopulations of social scientists) treat culture as a "get out of jail free" card, one that instantly dissolves the obligation to reconcile their accounts of altruism with the constraints of human biology and Darwinian evolution. In reality, the relationship between culture, biology, and natural selection is much more complex. In order for culture and biology to diverge in interesting ways — and thus for culture to provide an independent platform for the emergence of altruistic behaviour — there must be some *bias* in the way the cultural variants are acquired.

Perhaps unsurprisingly, there are such biases in the way that human culture is transmitted. In particular, when deciding what or who to imitate, we appear

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

to employ a set of heuristics that use indirect cues as a way of selecting the most *successful* variant in the population. One of the cues that we rely upon is simply frequency, such that the behaviour most likely to be imitated is that which is the most *common* in the population to which an individual is exposed.[41] It is not difficult to see how such a conformist bias could be (genetically) adaptive. Since the most successful variant is likely to become the most common, one can increase the odds of acquiring it by switching from an unbiased sampling procedure to one that is frequency-dependent. This allows a population of social learners to converge much more quickly upon the most adaptive phenotype. But it also has interesting side effects. The most significant, for our purposes, is that it greatly amplifies the power of group selection, and thus may indirectly promote the evolution of altruistic behaviour.[42] It does so by creating tipping-point effects, within groups, that greatly increase the level of correlation at the level of phenotype (and thus increases the likelihood that altruists will interact with altruists, and selfish agents with selfish agents). For this reason, Boyd and Richerson describe culture as "potentiating" group selection.

These are the three different mechanisms that have been posited as explanations for altruism in nature. However, it should be noted that the third — cultural transmission — has been posited first and foremost as a way of explaining the extremely high rates of altruistic conduct among human beings. *Homo sapiens* are one of a very small number of so-called ultrasocial or hyperco-operative species. We sustain and reproduce life through unusually extensive and complex systems of interdependence. In fact, human beings have recently begun to overtake ants, termites, and bees as the most social species on the planet (in terms of the number of individuals and the degree of complexity of our societies).[43] This is uncommon. But apart from being uncommon, there is also something unusual about the specific form that ultrasociality takes on in human societies. In the cases of both colonial invertebrates and social insects, there appears to be a genetic basis for the peculiarly high levels of sociality that they exhibit (viz. "haplodiploidy," although the precise significance of this is subject to dispute).[44] There is nothing comparable in the reproductive biology of humans. With respect to reproduction, and in many other aspects, we are not all that different from apes. Yet apes, who make significant parental investment in their young and exhibit a sophisticated capacity for reciprocal altruism, tend to congregate in tribes of no more than one hundred individuals, without any large-scale co-operation. The fact that they are able to sustain such weak levels of social cohesion suggests that inclusive fitness and reciprocal altruism are not especially robust mechanisms. And so, despite the rather optimistic pronouncements of sociobiologists over the past three decades, it seems increasingly improbable that these two mechanisms are sufficient to explain the level of co-operation one finds in human societies.[45] If they were, then we would expect to see a lot more altruism, and a lot more social structure, among our nearest primate relatives.[46]

Thus we are faced with something of an evolutionary puzzle. Richerson and Boyd describe it this way:

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

> Humans are, arguably, a new page in the natural history of animal co-operation. Our reproductive biology is similar to the other social mammals. Among our close relatives, the apes and monkeys, genetic relatedness and reciprocal altruism support a diverse array of small-scale societies, but no other spectacular ones. Humans have built extremely complex societies by some mechanism or mechanisms different from any other known highly social species. At the same time, there are remarkable parallels between human and ape social behaviour and material culture, not to mention many convergences between humans and other social and tool-using species. Consistent with classical comparative anatomy and modern molecular studies, human behaviour is clearly recently derived from ape behaviour. There is room for only relatively few modifications of the behaviour of the last common ancestor of chimpanzees and humans.[47]

It should be noted that cultural transmission is not the only mechanism that has been posited to explain human ultrasociality. Furthermore, it is something of an umbrella concept. Once cultural transmission has been posited, there is still room for a wide range of competing hypotheses about the mechanism that might sustain altruism in the cultural domain (e.g., altruistic punishment, strong reciprocity, group selection, etc.). I will not have an opportunity to discuss these rival claims. For the purposes of this paper, I will focus upon the specific proposal made by Boyd and Richerson — which focuses on conformist imitation as a mechanism of transmission — in order to illustrate the impact that it would have upon our thinking about morality.

Altruism and the Adaptive Unconscious

My primary hypothesis in this paper is that each of these three mechanisms has correlates in our adaptive unconscious — cognitive modules or heuristics that are used to analyze social situations and to generate altruistic behavioural dispositions when triggered. Because cognitive processing done at this level is fast, dedicated, and unconscious, such altruistic dispositions will tend to be associated with spontaneous, largely involuntary responses that we have to social situations. Thus the evidence for the hypothesis lies in the "fit" that one can observe between these responses and the different evolutionary mechanisms outlined above.

1. Sympathy

Given the extremely high level of parental investment required to raise human infants to maturity, one would expect humans to be powerfully affected by altruistic impulses that enhance inclusive fitness. In order to see this, one need only observe the sort of behavioural repertoires that are typically evoked in the presence of infants. In many mammalian species, infants are characterized by a set of highly distinctive "neotenous" or juvenile characteristics. This includes fur of a different colour, distinctively proportioned facial and body features (e.g., large eyes, small nose, large head-to-body size ratio), and "a staggering

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

'infantile' gait."[48] These characteristics in turn bring out distinctive responses in adults within many species, ranging from caring behaviour and protection, a suspension of aggressive behaviour, and much greater tolerance of impropriety in social interaction (e.g., violations of personal space, failure to respect the dominance hierarchy, etc).[49]

Human response to neotenous characteristics — or "cuteness," as it is referred to colloquially — follows a very similar pattern. There is clearly an important element of visual preprocessing. Certain faces simply "look cute," others do not, and most of us have no idea what makes them so. It turns out that the cutest tend to be those that are supernormally neotenous, i.e., possess characteristically neotenous features to the highest degree.[50] Furthermore, adult response to infantile facial characteristics is largely invariant across persons and cultures.[51] Other aspects of the way adults conduct themselves in the presence of infants are culturally invariant as well. For example, most adults spontaneously adopt a modified form of speech known as "baby talk" when addressing an infant. Despite enormous differences in the sound patterns of different languages around the world, "baby talk" has a similar intonation pattern.[52] Thus it is widely believed to be an altruistic adaptation, designed to make language more learnable — or at least verbal communication more intelligible — for the young.

Examples such as these suggest that the way adults identify neoteny and are disposed to respond to such characteristics is the product of a set of evolutionarily adapted cognitive mechanisms selected for reasons of inclusive fitness. The "cute" reaction is, of course, quite general, evoked not only by most human infants but by the infants of many other mammalian species. Thus there may be some question as to how such a non-discriminating mechanism could evolve through the kin mechanism. There are several possible responses: first, the extent to which it is shared with other species suggests that it is extremely archaic, and thus predates the development of more sophisticated discriminatory abilities; second, individuals are more likely to be in the company of their own infants, and so the latter are most likely to be the ones who benefit from even a perfectly general disposition (in the same way that anyone can hear an alarm call, and yet in many species these calls are thought to have evolved through kin selection);[53] finally, there is evidence that parents are inclined to regard their own children as more attractive than those of others, and that the identification of "family resemblance" is an important aspect of parental bonding.[54] This suggests that there may be some element of kin discrimination built into the mechanism.

One can see a similar structure in the level of sympathy that individuals experience for others, particularly when those others are in distress. It is well-known that the intensity of our experience of sympathy seems quite closely tied to our capacity for identification with the one who suffers, which is in turn closely tied to an experience of similarity. Furthermore, sympathy is more directly evoked by visual observation than through any other mode of encounter. Since those who resemble us most are our relatives, this structure bears all the

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

hallmarks of a psychological disposition adapted to generate altruistic behaviour of primary benefit to immediate kin. Finally, there is a significant literature on the impact that neotenous appearance has on criminal proceedings, which suggests that while it has a relatively minor impact on rates of conviction, its does promote much more lenient sentencing.[55] This certainly coincides with the impression many people have that those who are more attractive, or more neotenous in appearance (the two measures largely coincide[56]), are often treated with greater sympathy by others.

These limitations on the *scope* of sympathy all point in the direction of kin selection as an adaptive explanation. As David Hume famously observed, "[A] man naturally loves his children better than his nephews, his nephews better than his cousins, his cousins better than strangers, where every thing else is equal."[57] This is close to being a precise statement of Hamilton's calculus of inclusive fitness. There are other features of the sympathy response that also suggest an explanation in terms of inclusive fitness. First and foremost, there is the fact that it is a largely involuntary and unconscious response, usually narrowly limited in scope, and mediated through a mechanism of identification such as visual similarity. Second, a basis in kin selection suggests that it may rely upon very archaic psychological structures, which would explain why even very young children exhibit spontaneous empathy or caring behaviour prior to the development of more advanced cognitive skills such as instrumental reasoning or role-taking. Finally, it would explain why there are (small, but statistically significant) sex differences in the nature and intensity of these reactions — given that the optimal level of parental investment is unlikely to be the same for males and females.[58] (It was presumably casual familiarity with these sex differences that led many philosophers to accept the assertion that women were more likely to adopt a "care" orientation in their moral reasoning, despite the absence of scientific evidence in support of this claim.[59] When methodologically sound investigations were conducted, there turned out to be no significant differences between men and women at the level of intentional moral reasoning.[60] This is precisely what one would expect, given the language-dependence of the intentional planning system. Shared structures of moral reasoning are perfectly compatible with significant sex differences at the level of the adaptive unconscious, just as the same software can run on different underlying hardware. Thus women may be more inclined to engage in caring behaviour, and may exhibit more spontaneous sympathy, but this does not imply that they will reason any differently when confronted with moral problems.)

2. Virtue

By comparison to inclusive fitness, reciprocal altruism is a mechanism that requires a much higher level of cognitive sophistication to develop. In particular, it requires the capacity to distinguish those who have co-operated in the past from those who have not, along with the ability in many cases to keep a "running tab" on how balanced the co-operative exchange has been. One of the primary

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

functions of reciprocal altruism in our closest primate relatives is to support the formation of coalitions and alliances. Among chimpanzees, bonobos, and some baboon species, an "animal's chances of securing food, receiving care and protection, exercising mate choice, and other important determinants of fitness depend on which other members of the troop will come to their aid."[61] Furthermore, there is evidence that in some species, including chimpanzees, the existence of an altruistic "relationship" in one category of service increases the likelihood that others will be established, and even that some forms of altruistic service will be exchanged against others. This explains the asymmetries that sometimes exist in the exchange of grooming services. For example, chimpanzees are more likely to share food with those who have groomed them in the past.[62] Experimental studies have also suggested that in agonistic interactions, ververts and macaques are more likely to intervene in support of individuals who have groomed them in the past.[63]

In other words, the types of relationships developed through reciprocal altruism bear a close resemblance to what we call "friendship." Many theorists have speculated that the seemingly universal human propensity to categorize individuals that we meet into "friend" and "foe" (or those whom we "like" and "dislike") springs from a similar coalition-forming disposition.[64] It could also explain the cognitive bias that social psychologists refer to as the "fundamental attribution error," whereby people systematically overestimate the importance of character traits or "virtues" in explaining the behaviours of others, and underestimate the importance of situational factors. The psychological mechanism would be quite simple. Benefiting from an altruistic act performed by another generates a slight positive cathexis of that individual (one begins to like him, or to ascribe a "virtuous" character). Suffering from a selfish or aggressive act performed by another generates a negative cathexis (one begins to dislike him, or ascribe a "vicious" character). This is how we passively keep track of the benefits we are receiving. Liking someone in turn generates a highly generalized disposition to act altruistically toward him or her.[65] (Our dispositions follow the well-known pattern of helping our friends and harming our enemies.) The reason that we are inclined to think about action in terms of character rather than situation is precisely that this mechanism requires that we keep score on, and thus credit the account of, individuals, not situations.

In this context, it it important to remain clear on the distinction between conditional co-operation as a rational strategy and the type of dispositions produced by evolution. The latter may generate genuine, unconditional altruism. The function of reciprocity is simply to shield this disposition from being eliminated by the forces of natural selection. Thus the development of feelings of friendship toward those who have helped us in the past may create on our part a genuine disposition to act altruistically toward them (based perhaps upon a "warm" feeling that we have when thinking of them), even if at the intentional level we come to anticipate that this action will go unreciprocated. Indeed, some have observed that when individuals explicitly keep track of the "favours"

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

that others have done for them, it tends to undermine the emergence of friendly feelings.[66] This does not imply, however, that friendship is not a product of reciprocal altruism. What it suggests is that the emergence of friendship relations occurs at the level of the adaptive unconscious, not the intentional planning system. As Frans de Waal writes:

> Whether what is involved is the returning of a favour or the seeking of revenge, the principle remains one of exchange; and, most importantly, this principle requires that social interactions be remembered. Much of the time the process may take place in the subconscious, but we all know from experience that things come bubbling up to the surface when the difference between costs and benefits becomes too great. It is then that we voice our feelings. By and large, however, reciprocity is something that takes place silently.[67]

3. Imitative conformity

The most important psychological feature of humans associated with culture-dependence is the fact that we are extraordinarily imitative. (Psychologist Michael Tomasello has described human infants as "imitation machines.")[68] Although social learning is relatively common in nature, genuine imitativeness is not. Most animals do not directly copy the behaviours of conspecifics, but rather make inferences based upon cues in the social environment.[69] Chimpanzees, for instance, appear to engage in what Tomasello calls "emulative learning," which differs from the sort of imitation seen among human infants. He and his colleagues conducted an experiment in which they demonstrated, for both chimpanzees and human infants, the use of a tool to obtain an out-of-reach object.[70] The tool could be used in one of two ways, with one method being clearly more effective than the other. Both human infants and chimpanzees learned to use the tool but the chimpanzees tended to use the tool in either of the two ways, regardless of which method had been demonstrated. Human infants, on the other hand, used the tool in precisely the fashion that had been demonstrated.[71] Thus what chimpanzees learned from the demonstration was simply *that* the tool could be used to obtain the food. They used their own intelligence in order to determine how to use it (and in this sense, reinvented the wheel). Human infants, on the other hand, learned not simply *that* the tool could be used, but *how* it was used as well. They imitated the *actions* of the other, even if their own instrumental reasoning might have led them to a better solution.

This imitative propensity persists throughout adulthood. It is used, among other things, to establish what Ap Dijksterhuis and John Bargh refer to as "default social behaviour."[72] Furthermore, rather than being strategic or intentional, it is almost entirely automatic and unintentional.[73] Dijksterhuis posits that "We do not choose to imitate at some times and not others. Rather, we are wired to imitate and we do it all the time, except when other psychological processes inhibit imitation."[74] The reason it is more noticeable in infants is that they have not yet

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

developed higher-order mental or cognitive processes able to "trump" it. Yet even among adults, the propensity is surprisingly pervasive, especially in the way that it biases high-order cognition and intentional behaviour. As Dijksterhaus argues, imitation is exhibited "in a very broad domain of behaviour":

> We adjust motor behaviour and a range of interpersonal behaviours such as helpfulness or aggression; it affects mental performance in different ways; and it affects our attitudes. More concretely, relevant research has shown by now that imitation can make us slow, fast, smart, stupid, good at math, bad at math, helpful, rude, polite, long-winded, hostile, aggressive, co-operative, competitive, conforming, nonconforming, conservative, forgetful, careful, careless, neat, and sloppy.[75]

As infants become older, the fact that imitation exhibits a conformist bias also become more apparent. This is observable primarily in the way that peer-group norms take over from parental cues as the most dominant model for social behaviour. For example, grade-school children quickly lose whatever spoken accent they may have acquired in the home and adopt forms of pronunciation that are close to the mode of the broader group of speakers to whom they are exposed. Much of this peer influence is exercised not through sanctions or explicit prescription, but simply by triggering the impulse to engage in conformist imitation.[76]

Of course, depending upon what others are doing, conformity or imitation may promote outcomes that are either altruistic or non-altruistic. Thus the underlying psychological disposition differs from both sympathy and friendship because it does not intrinsically orient the subject toward more pro-social behaviour at the level of outcome. The reason that imitative conformity is important, from the standpoint of altruism, is that it introduces a significant *non-teleological* element into human motivation. This in turn removes one of the most important filters than constrains the emergence and reproduction of altruistic behaviour patterns at the population level, viz. the possibility that these patterns will be undermined by individual intelligence and teleological action. In the same way that a clever chimpanzee (using an emulative learning strategy) is more likely than a human infant (using an imitative learning strategy) to substitute the more effective method of tool use for the one that was demonstrated, the same chimpanzee is also more likely than the human infant to adopt a free rider strategy, having observed a co-operative interaction. In this case, the problem with chimpanzees is precisely that they think for themselves a little too much.

Another way of putting it is to say that unlike sympathy and friendship, which are mechanisms that encourage individuals to care more about the interests of others, imitation promotes altruism only indirectly, by making individuals care less about their own interests (an effect that is achieved by making them less concerned with *anyone's* interests, insofar as these are associated with outcomes achieved). Thus conformist imitation is important, not because

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

it directly promotes altruism in individual actions but because it facilitates the reproduction of altruistic behavioural patterns at the level of culture.

What this Means for Us

The discussion in the previous section has consisted of fairly speculative reflections about what are essentially empirical questions. The motivation for these reflections, however, lies in the implications that they have for our understanding of morality. Thus I would like to end by discussing what I take to be the normative import of these empirical issues. Suppose it is true that we have three rather different sets of evolutionarily adapted cognitive heuristics, each structured in a way that generates altruistic behavioural dispositions of three fundamentally different types. Clearly this means that there is no unified altruistic psychology at the level of the adaptive unconscious. Insofar as we are naturally disposed to care about the welfare of others, we have a variety of different motives for doing so, and these impulses will be triggered in very different contexts. It is important to emphasize, however, that the dispositions generated are only that; dispositions; they are not desires (or full-blown intentional states). Thus the analysis being offered here is very far from being a sociobiological account of altruistic action, much less morality proper. Indeed, because I do not regard any specific cognitive processes occurring within the intentional planning system as exhibiting direct adaptation, it is an open question what we choose to do with our altruistic dispositions at that level. What reasons do we have for turning an altruistic impulse into a pro-social preference?

An attractive framework for addressing this question is the "sentimental rules" account of moral judgment proposed by Shaun Nichols. Nichols argues that "moral judgment depends on two dissociable mechanisms: an affective system and a normative theory."[77] The affective component constitutes the behavioural dispositions that are the output of a set of evolutionarily adapted heuristics. The normative theory, on the other hand, serves as the basis for the explicit component of moral judgment and is derived largely from a set of culturally transmitted norms that have been internalized by the individual. In principle the two components could produce quite different verdicts; indeed, the two systems could diverge completely. In practice, however, they do not, because the adapted heuristics bias cultural reproduction in a way that favours "norms prohibiting actions that elicit negative affect."[78] This is because our adapted psychology constitutes an important component of the *environment* within which the cultural evolution of norms occurs, and with respect to which the "fitness" of these norms is determined.

In Nichols's view, standard sentimentalism mistakes an empirical correlation for a necessary connection. Thanks to the adaptive unconscious, we are disposed to perform certain sorts of actions and to refrain from others. This is not the basis, however, for our explicit judgment that they are right or wrong. The judgment is merely correlated with our sentiments; it is not based on them. What it is *based* upon is the conformity of the action to the norms that we endorse, which

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

are typically derived from those that prevail in our culture. The fact that certain norms prevail in our culture, however, is influenced by the fact that we are disposed to perform certain actions or refrain from them. Norms that run against the grain of our evolved dispositions are at a reproductive disadvantage compared to norms that generate less tension, and so tend not to occur in our normative theory.

One of the attractions of this account is that it explains why we experience morality as a unitary phenomenon, despite the fact that at the level of the adaptive unconscious it may lack coherence. Jonathan Haidt and Fredrik Bjorklund speak for many evolutionary psychologists when they write that "The history of moral psychology is full of failed attempts to derive all of morality from a single source (e.g. noncontradiction, harm, empathy, or internalization). We think it makes more sense to look at morality as a set of multiple concerns about social life, each one with its own evolutionary history and psychological mechanism. There is not likely to be one unified moral module, or moral organ."[79] For example, the Westermarck effect, which many think underlies the prevalence of the incest taboo in human cultures, is pretty much a self-standing mechanism.[80] It is also not an altruistic disposition, and so requires no special evolutionary structure to sustain it. In this respect it has almost nothing in common with the psychological mechanisms that might reasonably be thought to promote, say, parental investment in the young. And yet given that this is so, it is not obvious why we consider both parental abandonment and incest to be "moral" issues. The explanation proposed by Nichols is that they both involve norm-violations — in our culture — and the norms in question have strong affective resonance. Thus they get assigned to the same category, despite the fact that the underlying affect differs in the two cases.

The other attractive feature of this account is that it opens up a certain amount of space between our evolved dispositions and our moral judgments. It is therefore able to accommodate in a fairly natural way both the fact that moral norms vary enormously from culture to culture and the fact that individuals are able to override particular altruistic dispositions when they choose to do so. It is the latter phenomenon that blocks any straightforward *identification* of altruistic behavioural dispositions and moral judgments. While we often choose to go with our intuitive judgments and dispositions, we also are capable of ignoring them. Animal rights activists, for instance, frequently publicize their cause using photographs of especially neotenous juveniles of other species (e.g., harp seals) in an attempt to expand our circle of moral concern. In the same way, anti-abortion activists use photographs of late-term foetuses (e.g., sucking on their thumbs) to achieve the same effect. Yet the intersection set of animal rights activists and pro-lifers is very small, suggesting that almost everyone, in certain cases, is capable of overriding the altruistic impulses that arise from their responses to neoteny.

Yet while the sentimental rules account opens up enough space to permit critical reflection and revision of the behavioural dispositions that arise at the

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

level of the adaptive unconscious, it also brings to the fore an obvious skeptical problem. If we are capable of ignoring some of these behavioural dispositions, how are we to decide when we should and should not do so? There are many cases, for instance, where altruistic behavioural dispositions conflict with moral norms that we endorse. Retributivist urges, for instance, which are no doubt part of a behavioural complex that sustains reciprocal altruism, must often be suppressed in order to make way for more inclusive forms of co-operation. Indeed, the tit-for-tat structure of strong reciprocity is capable of generating highly dysfunctional interaction patterns, such as vicious cycles of violence and retribution (e.g., blood feuds). The willingness of individuals to abstain from the private use of force and cede to the state the right to punish those who violate the normative order is an important step in the advancement of civilization. The fact that it runs against the grain is reflected in the fact that many people find this arrangement a source of enormous frustration.

Similarly, the mechanisms that promote parental investment and care for the young appear to have features designed to *limit* parental investment in cases where doing so may be unreasonably costly or the offspring is unlikely to be viable. The neoteny response, for instance, is partially vitiated by the presence of craniofacial abnormalities, many of which are associated with chromosomal or developmental damage. There are a variety of other circumstances that appear to interfere with parental bonding, including maternal stress. Setting aside the question of postpartum depression, impaired bonding affects a small but significant number of otherwise normal mothers, many of whom are extremely distressed by the lack of affection — indeed, sometimes the revulsion — they feel for their infant.[81] It is not uncommon for these mothers to experience unwanted yet persistent infanticidal ideation.[82] Of course, the fact that these thoughts are unwanted is at least in part a consequence of the fact that infanticide is subject to a strong moral prohibition in our culture. Yet historically this has not been the most common arrangement. Infanticide has typically been tolerated and in many instances (such as Roman law) made mandatory in cases of obvious deformation. Although the force of our own taboo makes it difficult to contemplate the matter dispassionately, it seems fairly likely that these earlier norms had positive affective resonance and that in our own society, when circumstances trigger such dispositions the norm against infanticide is upheld against the grain.

Additional problems are introduced by the fact that the behavioural dispositions we are dealing with are typically based on heuristics, which means that they often are not directly associated with the favoured outcome but rather with some proxy or stand-in, something that was strongly correlated with it in the environment of evolutionary adaptation. This can generate a certain amount of misfiring, so that even when we endorse the overall biological "objective," we can see that our own dispositions are failing to achieve it. With the Westermarck effect, for instance, the "purpose" was presumably to reduce the incidence of offspring with homozygous pairings of genetic disorders. What the disposition

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

targets, however, is quite removed from that. It appears to generate an aversion to sexual relations amongst those who were in close contact over an extended period of time during childhood. This is in turn a proxy for relatedness, which is in turn a proxy for genetic similarity. Furthermore, the objective is to avoid not sexual relations, but rather reproduction. As a result, some people (including some moral philosophers[83]) think the incest taboo needs to be relaxed (with perhaps emphasis put instead on contraception and genetic screening).

One of the more common types of misfires, however, involves excessive generosity in our altruistic dispositions, which in turn makes people vulnerable to exploitation. It is difficult to know how we should feel in these cases. One naturally feels a sense of outrage when one sees pictures of magpies working desperately to feed gigantic baby cuckoos. This is an obvious case of parasitism, since the desire to feed and care for the young, having evolved for reasons of inclusive fitness, was obviously never intended to benefit members of another species. And yet if one looks at the relationships between humans and their pets, certain uncomfortable similarities can be observed. Several species of domestic animal have evolved in ways that allow them to better exploit the ecological niche provided by human emotional responses. Dogs in particular have evolved in the direction of increased neotenousness by preserving a variety of juvenile characteristics (such as barking and playfulness) into adulthood, partly because it facilitates domestication but no doubt partly also because it triggers certain human emotional responses.

Of course, once we start to ask too many questions about the biological "purpose" of particular behavioural dispositions, it raises the obvious question of why we should care about any of them at all.[84] The response we have to the neotenous appearance of a puppy is clearly a misfire, from an evolutionary perspective. But why should we assign any more authority to the response we have to the neotenous appearance of one's own child? Looking at the whole thing from an evolutionary perspective naturally encourages the radical thought that we might choose to regard all of our altruistic dispositions as simply a ruse being perpetrated upon us by us selfish genes. (Why is the sight of a magpie working hard to feed a cuckoo any more pathetic than the sight of a magpie working hard to feed a baby magpie?) Indeed, the moment that organisms acquire the sophistication required to learn new forms of behaviour, they often have an incentive to "unlearn" many altruistic behaviour patterns, especially those that arise for reasons of inclusive fitness. For example, chimpanzees produce a characteristic vocalization (or "call") when they discover a food source. This behaviour is potentially altruistic, since it attracts the attention of others and may result in the individual who made the discovery being denied access to the food by those further up the dominance hierarchy. Chimpanzees are smart enough to realize this, but they lack the cortical brain structures required to exercise direct control over these vocalizations. Thus Jane Goodal observed one chimpanzee, who had recently come across a cache of bananas at the Gombe reserve, emitting the standard vocalization associated with food discovery while at the

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

same time putting his hand over his mouth in an effort to muffle his own cry.[85] We can see here an early stage what Keith Stanovich calls "the robot's rebellion" against the selfish gene.[86] Just like the science fiction stories in which intelligent machines one day achieve self-consciousness and begin to reject their enslavement to humanity, we are essentially machines built by our genes who have only recently achieved awareness of our origins. It is only natural that we should begin to question why we are continuing to follow our programming, in cases where doing so involves sacrificing ourselves for the benefit of our creators.

So one can see three distinct sets of considerations that might dissuade us from acting upon our altruistic dispositions, or transforming them into prosocial preferences. What sort of philosophical response is there to this problem? This is of course not the place to provide a comprehensive solution to the problem of moral skepticism. There are, however, three well-established philosophical strategies for addressing this issue, which I refer to, for convenience, as the Humean, Millian, and Kantian responses. What I would like to show, in the remainder of this section, is that the more pluralistic altruistic psychology outlined above provides resources for amplifying these three traditional argumentation strategies, making each one more persuasive than it was in its original formulation.

The Humean response to moral skepticism has two prongs, corresponding to Hume's distinction between natural and artificial virtue.[87] The first argument involves simply pointing out that our "sentiments" are relatively unaffected by philosophical reflection, so that even after we come to the realization that the affection we feel for our children (for example) is an elaborate ruse, it doesn't really change the way we feel about them. Since most people are this way, we don't have to worry too much about the possibility of a "robot's rebellion," or a large-scale rejection of natural virtue. Elaborating somewhat, we can say that since we are "gene-built" from the ground up, it is not clear how we could develop a conception of "our" interests that would be independent of our genes' interests. After all, people derive great satisfaction from caring for, raising, and even making sacrifices on behalf of their children. The idea that there is some perspective from which it can be decided that this is contrary to one's interests seems entirely illusory. Thus there is good reason to expect the preponderance of people simply to go with the feelings and behavioural dispositions that they have.

The second aspect of Hume's argument involves the observation that certain altruistic behavioural dispositions can fairly easily be channelled into explicit preferences that will support a mutually beneficial system of co-operation (promise-keeping, respect for property, etc).[88] Furthermore, many of these systems of co-operation are formally institutionalized and reinforced through sanctions. Since we derive benefits from our participation in these institutions, and since failure to exhibit the relevant sort of preferences can lead to exclusion or punishment, it is simply not in our interest to strip ourselves of many altruistic behavioural dispositions. Thus there are good prudential reasons for going along with many of our altruistic dispositions, since doing so allows us to remain in

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

good standing with others, which in turn gives us access to socially maintained systems of reciprocity.

Neither of these two arguments is particularly powerful. The first is not intended to be. The second falls short because of Hume's tendency to overestimate the robustness of reciprocity as a mechanism for sustaining co-operation. From a strictly prudential perspective, free riding has a lot more to recommend it than many people suspect. The argument can be bolstered, however, by John Stuart Mill's observation that not only do we have an interest in adopting pro-social preferences, we also have a powerful incentive to persuade other people to adopt more pro-social preferences, in the forms of desires or principles that are likely to induce co-operative action. One of the ways that we do so is by seizing upon whatever altruistic dispositions are present in others and using them to encourage the adoption of preferences that will benefit ourselves and others. This involves *inter alia* structuring the deliberative field in such a way as to expand, and in some cases to redirect, basic altruistic impulses. Mill argues, with respect to sympathy, that "whatever amount of this feeling a person has, he is urged by the strongest motives both of interest and of sympathy to demonstrate it, and to the utmost of his power to encourage it in others; and even if he has none of it himself, he is as greatly interested as anyone else that others should have it."[89] Thus pro-social preferences will tend to be common, because even those who choose to suppress their own altruistic dispositions nevertheless have an incentive to cultivate them in others (which may be why hypocrisy is more common than frank selfishness).

The force of Mill's observation is somewhat diminished by his assumption that sympathy is the only "natural sentiment" that can be leveraged into more general moral motivation. The evolutionary analysis presented here suggests that additional resources are available. Not only do we attempt to extend sympathy in order to generate a more general notion of care and concern for others, we also try to take basic dispositions associated with friendship and use them to promote teamwork and group solidarity, or to build up a notion of shared citizenship.[90] We also take conformist imitation and use it to promote respect for various social norms, as well as "golden rule" reasoning in cases where settled norms are absent or indeterminate.

More speculatively, I would argue that the attempt to leverage our altruistic dispositions into pro-social intentional states is what underlies the three "moral psychologies" that have structured the history of moral philosophy. Appealing to sympathy is a way of convincing individuals that the suffering of others should serve as a relevant consideration when deciding what to do. Appealing to virtue and "the good" is a way of persuading them that *shared* goals should be assigned priority over self-interest, narrowly construed. And finally, appealing to duty is a way of convincing individuals that the way they go about achieving their objectives should be constrained by the prevailing set of social norms. (In this respect, the fact that the Kantian tradition in ethics emphasizes purely formal constraints on action is a reflection of the fact that the underlying

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

motivational disposition — what Kant referred to as *Achtung fürs Gesetz*, but referred to here as conformist imitation — itself has a purely formal structure; it does not directly impose concern for the interests of others.) All of these mechanisms limit the extent of collective action problems, the first two by reducing goal incongruity between agents, the final one by discouraging individuals from making exceptions of themselves.

This is of course a sociological observation, not a philosophical response to moral skepticism (since a clever free rider can do better by cultivating such expansive desires in others, without adopting them himself). Thus there is a final argument that can be made with respect to the possibility of a "robot's rebellion," one that is broadly Kantian in inspiration. While it may be possible to ignore some altruistic impulses some of the time, it may not be possible to ignore all of them all of the time. This is not just an empirical question; there is also an important conceptual question that arises, particularly with any proposed attempt to refrain from conformist imitation. One can at least imagine a person learning to suppress any feelings of sympathy or friendship for others (perhaps through some combination of will power and psychological conditioning). The idea that one might refrain from conformist imitation, however, encounters important conceptual difficulties. This is because the capacity for imitativeness among humans appears to be one of the building blocks that permits the emergence of a number of higher-order cognitive capacities, including intentional planning. There are a variety of ways of developing this idea, but those with pragmatist temperaments will be impressed by the role that imitative conformity plays in generating what Brandom calls "norms implicit in practice," which are in turn thought to be essential to the emergence of linguistic communication.[91] Imitative conformity, according to this view, is the competence that permits the emergence of rule-following. This rule-following competence is exhibited not just when individuals conform to social norms, but also when they participate in "the game of giving and asking for reasons," which is the social practice underlying rational deliberation. Thus practical deliberation at the level of the intentional planning system is fundamentally an exercise in rule-following, which requires imitative conformity. This means that there is something fundamentally incoherent about deciding, on the basis of such deliberation, to discontinue one's habit of following rules. How is one expected to detach the conclusion arrived at through such a chain of reasoning?[92]

If this could be shown successfully, then the rudiments of a transcendental justification of our willingness to be motivated by a conception of duty would have been offered (providing a somewhat unexpected validation of the fundamental Kantian intuition that there is an internal connection between acting rationally and acting from a motive of duty). Not only would it be empirically demonstrable that our intentional planning is shot through with preferences that arise out of our underlying disposition toward conformist imitation, but we would have a philosophical argument demonstrating the rational incoherence of trying to purge ourselves of such preferences. Of course there would still be

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

important questions about the content of the norms that we should be following, such as what we should say about misfires. What Christine Korsgaard calls "content skepticism" with regard to morality would still be a problem.[93] However, the problem that Korsgaard refers to as "motivational skepticism" will have been taken off the table. A robot's rebellion will have been shown to be impossible, simply because some of our higher cognitive competences turn out be byproducts of the adaptation that is responsible for much of our pro-social behaviour.

Conclusion

The contribution made by evolutionary theorists to debates in moral philosophy have, so far at least, not been especially useful. Setting aside those who get into naturalistic fallacy trouble or who choose to engage in jejune "debunking" of moral concepts, even those who have attempted a friendly engagement between fields of inquiry have seldom done justice to either. The most common difficulty arises from the assumption that some standard Darwinian model of evolutionary selection acts directly upon either the content of the intentional states of individuals or the actions that they choose.[94] This typically involves a rather studied disregard for one of the features of human moral psychology that has been a particular preoccupation for philosophers, viz. our capacity to rationally reflect upon and possibly revise our desires and intentions. For example, despite all the ink that has been spilled over the tit-for-tat strategy in evolutionary game theory, it is seldom mentioned that tit-for-tat is not a subgame-perfect strategy in a rationality-based repeated prisoner's dilemma game.[95] Any player smart enough to anticipate more than one move ahead will have an incentive to deviate from it (since the consequences of punishing someone are identical to the consequences of initial defection). This means that in order to be willing to play tit-for-tat, individuals have to be either hard-wired to play it or else completely myopic. As a result, tit-for-tat models tell us *absolutely nothing* about how or why rational agents might choose to co-operate with one another. Furthermore, any suggestion that they do implicitly denies not just the human capacity for rational reflection and preference-revision, but even our "bare bones" capacity for instrumental reasoning.

Shifting away from a sociobiological framework toward a "dual inheritance" or "gene-culture coevolution" model does not guarantee a more productive engagement with moral philosophy, either. Standard models in the field of "memetics" tend to regard individuals as what Harold Garfinkel once called "cultural dopes" — individuals blindly acting out programs imposed upon them by their memes rather than by their genes. Substituting cultural evolution for biological evolution does not achieve very much, if the framework is one that continues to ignore the role that individual rationality plays in either filtering or propagating cultural patterns. Furthermore, crude memetic models can lead to an unhelpful occlusion of biological influences, insofar as they ignore the way that adapted psychological propensities can powerfully bias the transmission of cultural patterns.

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

The way to make it all fit together in a more useful way, I have tried to show, is by combining a dual inheritance model with a dual-process theory of mind. In this way, biological mechanisms can be analyzed in terms of the ways that they affect modular processes at the level of the adaptive unconscious. The relationship between evolutionary biology and human action is then analyzed in terms of the relationship between the adaptive unconscious and the intentional planning system (with cultural influences remaining, at this point, an unexplored part of the story). The result is a framework that respects the voluntarism and the reflexivity of human action (i.e., the fact that people choose what to do, and that they typically know what they are doing, as they are doing it).[96] These are not only elementary postulates in any theory of rational action, they are also a fundamental aspect of our moral phenomenology.

The most useful feature of this framework, I would like to suggest, is that it allows us bring an evolutionary perspective to bear upon the question of where our inchoate moral "intuitions" come from. Moral philosophy, it has been widely noted, has come to depend upon such intuitions to an increasing degree in recent years (along with so-called "intuition pumps" — hypothetical scenarios that evoke intuitions with interesting structural features). Serious questions have been raised about the fruitfulness of such methods.[97] The analysis developed here suggests that, given the overlap that exists between altruism and morality, many of these intuitions are likely to be conflicting, not only between individuals but also within. Some people are likely to have highly deontological intuitions, others more axiological intuitions, and others are likely to find moral sentiment theory appealing. Furthermore, everyone is likely to be pulled in one or another of these directions on some occasions. No amount of philosophical analysis is likely to vindicate one set of intuitions at the expense of another. At the level of the adaptive unconscious, we simply do not have a unified moral psychology. Of course, this does not prevent us from adopting a relatively coherent and uniform moral code at the level of the intentional planning system, and of the culture more generally. But it does force us to recognize that we may, both as individuals and as a society, have to develop a number of work-arounds to the "bugs" that exist at the level of our altruistic dispositions.

Notes

1 Stephen Darwall, Allan Gibbard, and Peter Railton write: "Too many moral philosophers and commentators on moral philosophy — we do not exempt ourselves — have been content to invent their psychology and anthropology from scratch." In "Toward Fin de Siècle Ethics: Some Trends," *Philosophical Review* 101 (1992): 115–89, at 189.

2 Recent neurological research is certainly pointing in this direction. In the same way that "memory" has been shown to be not a single phenomenon but rather the joint product of several quite different systems, moral judgment also appears to be a neurologically complex phenomenon. See Joshua Greene and Jonathan Haidt, "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6 (2002): 517–23.

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

3 Keith Stanovich, *The Robot's Rebellion* (Chicago: University of Chicago Press, 2004), 34–6.
4 See Stanovich, 38. Also see Timothy D. Wilson, *Strangers to Ourselves* (Cambridge: Harvard University Press, 2006), 49. Note that this is not the Fodorian definition of modularity, and that the characteristics listed are only those typically associated with modular processes (they are not necessary and sufficient conditions). See H. Clark Barrett and Robert Kurzban, "Modularity in Cognition: Framing the Debate," *Psychological Review* 113 (2006): 628–47.
5 Robyn Carston, "The Architecture of the Mind: Modularity and Modularization," in David. W. Green, ed., *Cognitive Science* (Oxford: Blackwell, 1996), 73–5; Barrett and Kurzban, "Modularity in Cognition: Framing the Debate," 629–30.
6 Richard Samuels, "Is the Human Mind Massively Modular?" in Robert Stainton, ed., *Contemporary Debates in Cognitive Science* (Oxford: Blackwell, 2006), 48–52.
7 Wilson, 23–35.
8 Stanovich, 37.
9 Robert Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994), 5. I take the "intentional planning system" to be essentially equivalent, as well, to what Norman and Shallice refer to as the "supervisory attentional system," see Donald A. Norman and Tim Shallice, "Attention to Action: Willed and Automatic Control of Behaviour," in R. J. Davidson, G. E. Schwartz, and D. Shapiro, eds., *Consciousness and Self-Regulation, Vol. 4* (New York: Plenum, 1986).
10 Daniel Dennett, *Consciousness Explained* (Boston: Little, Brown and Company., 1991), 218.
11 Stanovich, 35–36.
12 Dennett, 218.
13 Dennett, 190. On exaptation, see Stephen J. Gould and Elisabeth S. Vrba, "Exaptation — A Missing Term in the Science of Form," *Paleobiology* 8 (1982): 4–15.
14 Joseph Heath, *Following the Rules* (New York: Oxford University Press, 2008).
15 Andy Clark, "Magic Words: How Language Augments Computation," in Peter Carruthers and Jill Boucher, eds., *Language And Thought* (Cambridge: Cambridge University Press, 1998), 164.
16 On language as the "vehicle of thought" see Michael Dummett, "Language and Communication," in *The Seas of Language* (Oxford: Clarendon, 1993), 166–87. I am, of course, implicitly rejecting the view that language is innate. I am not, however, denying the suggestion that language acquisition is facilitated by various innate learning heuristics, or that grammar is constrained by various innate structural features of the brain. These hypotheses often are not adequately distinguished in the literature.
17 Marc D. Hauser and Elizabeth Spelke, "Evolutionary and Developmental Foundations of Human Knowledge: A Case Study of Mathematics," in Michael Gazzaniga, *The Cognitive Neurosciences,* vol. 3 (Cambridge: MIT Press, 2004).
18 For an example of the "one more module" view, see Cosmides and Tooby, "The Psychological Foundations of Culture." In Jerome Barkow, Leda Cosmides and John Tooby, eds., *The Adapted Mind: Evolutionary Psychology and the Generation*

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

of Culture, (Oxford: Oxford University Press, 1992), 19–136.They describe the brain as "an intricate network of functionally dedicated computers, each activated by different classes of content or problem, with some more general-purpose computers embedded in the architecture as well," (p. 94). The somewhat "tacked on" feel of this final clause reflects the central weakness of the theory.

19 See Daniel Kahnemann, Paul Slovic, and Amos Tversky, eds., *Judgment Under Uncertainty* (Cambridge: Cambridge University Press, 1982). For a useful survey, see Stanovich, pp.110–13.

20 Amos Tversky and Daniel Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* 5 (1973): 207–32.

21 Peter J. Richerson and Robert Boyd, *Not by Genes Alone* (Chicago: University of Chicago Press, 2005), 127–9. See also Gary Marcus, *Kluge* (Boston: Houghton Mifflin, 2008), 161–4.

22 Stanovich, 131–46.

23 Elliot Sober and David Sloan Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Cambridge: Harvard University Press, 1998).

24 Anyone uncomfortable with this use of the term "morality" should feel free to use the term "pro-social action," instead. There are in fact certain advantages to doing so, since "morality" as the term is usually used includes a variety of norms, such as prohibitions on incest, that have no obvious benefit for others.

25 See Vero Copner Wynne-Edwards, *Animal Dispersion in Relation to Social Behavior* (Edinburgh: Oliver and Boyd, 1962).

26 Lee Dugatkin, *Cheating Monkeys and Citizen Bees* (New York: Free Press, 1998), 17–8.

27 George C. Williams, *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton: Princeton University Press, 1966).

28 Ronald Fisher, *The Genetical Theory of Natural Selection* (Oxford: Clarendon, 1930).

29 W. D. Hamilton, "The Evolution of Altruistic Behavior," *The American Naturalist* 97 (1963): 354–6; also J. Maynard Smith, "Group Selection and Kin Selection," *Nature* 201 (1964): 1145–7.

30 Richard Dawkins, *The Selfish Gene,* 2nd ed. (Oxford: Oxford University Press, 1989), 19.

31 Ibid., 2.

32 M. Soler, J. J Soler, J. G. Martínez, T. Pérez-Contreras and A. P. Møller, "Micro-Evolutionary Change and Population Dynamics of a Brood Parasite and Its Primary Host: The Intermittent Arms Race Hypothesis," *Oecologia* 117 (1998): 381–90.

33 Scott Woodcock and Joseph Heath, "The Robustness of Altruism as an Evolutionary Strategy," *Biology and Philosophy* 17 (2002): 567–90.

34 Robert L. Trivers, "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46 (1971): 35–57. For a useful recent survey, see Rajiv Sethi and E. Somanathan, "Understanding Reciprocity," *Journal of Economic Behavior and Organization* 50 (2003): 1–27.

35 Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984).

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

36 See Woodcock and Heath, also Robert Boyd and Peter Richerson, "The Evolution of Reciprocity in Sizable Groups," in *The Origin and Evolution of Cultures* (Oxford: Oxford University Press, 2005), 152–8.
37 Richard Alexander, *Darwinism and Human Affairs* (Seattle: University of Washington Press, 1979); also Richard Alexander, *The Biology of Moral Systems* (New York: Aldine de Gruyter, 1987); Martin A. Nowak and Karl Sigmund, "The Dynamics of Indirect Reciprocity," *Journal of Theoretical Biology* 134 (1998): 561–74.
38 See Olaf Leimar and Richard C. Connor, "By-product Benefits, Reciprocity, and Pseudoreciprocity in Mutualism," in Peter Hammerstein, ed., *Genetic and Cultural Evolution of Cooperation* (Cambridge: MIT Press, 2003), 203–22.
39 Robert Boyd and Peter J. Richerson, "Why Culture is Common but Cultural Evolution is Rare," *Proceedings of the British Academy* 88 (1996): 73–93.
40 Allan Rogers, "Does Biology Constrain Culture?" *American Anthropologist* 90 (1989): 819–31.
41 Richerson and Boyd, *Not By Genes Alone,* 120–4.
42 Joseph Soltis, Robert Boyd, and Peter J. Richerson, "Can Group-functional Behaviors Evolve by Cultural Group Selection? An Empirical Test," *Current Anthropology* 63 (1995): 473–94.
43 Peter Richerson and Robert Boyd, "The Evolution of Human Ultrasociality," in Irenaus Eibl-Eibesfeldt and Frank Kemp Salter, eds., *Indoctrinability, Ideology, and Warfare: Evolutionary Perspectives* (New York: Berghahn, 1998), 71–95.
44 E. O. Wilson, "Kin Selection as the Key to Altruism: Its Rise and Fall," *Social Research* 72 (2005): 159–168.
45 *Pace* Trivers, who in "The Evolution of Reciprocal Altruism" uses humans as his third example of reciprocal altruism.
46 For example, see the interesting study conducted by Felix Warneken and Michael Tomasello, "Altruistic Helping in Human Infants and Young Chimpanzees," *Science* 311 (March 3, 2006): 1301–3.
47 Richerson and Boyd, "The Evolution of Human Ultrasociality," 72.
48 Sarah Hall Sternglanz, James L. Gray, and Melvin Murakami, "Adult Preferences for Infantile Facial Features: An Ethological Approach," *Animal Behavior* 25 (1977): 108–15.
49 This hypothesis was first advanced by Konrad Lorenz. See C. F. Zachariah Boukydis, "Adult Perception of Infant Appearance: A Review," *Child Psychiatry and Human Development* 11 (1981): 241–54, at 245.
50 Boukydis, 247.
51 Sternglanz et al.
52 Anne Fernald, "Human Maternal Vocalizations to Infants as Biologically Relevant Signals: An Evolutionary Perspective," in Barkow, Cosmides, and Tooby, 398–401.
53 Dugatkin, 18.
54 Boukydis, 242.
55 See discussion in Leslie A. Zebrowitz and Susan M. McDonald, "The Impact of Litigants' Baby-Facedness and Attractiveness on Adjudications in Small Claims Court," *Law and Human Behavior* 15 (1991): 603–23.

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

56 Douglas M. Jones, "An Evolutionary Perspective on Physical Attractiveness," *Evolutionary Anthropology* 5 (1996): 97–109.

57 David Hume, *A Treatise of Human Nature*, 2nd ed., L. A. Selby-Bigge, ed., (Oxford: Clarendon, 1978), 483–4.

58 Martin L. Hoffman, "Sex Differences in Empathy and Related Behaviors," *Psychological Bulletin* 84 (1977): 712–22; Mark H. Davis, *Empathy: A Social Psychological Approach* (Colorado: Westview Press, 1994). There is also, unsurprisingly, evidence that women show a stronger response to babies and toddlers. See Katherine A. Hildebrandt and Hiram E. Fitzgerald, "Adults' Response to Infants Varying in Perceived Cuteness," *Behavioral Processes* 3 (1978): 159–72, at 169.

59 A claim first made by Carol Gilligan, *In a Different Voice* (Cambridge: Harvard University Press, 1982).

60 Lawrence J. Walker, "A Longitudinal Study of Moral Reasoning," *Child Development* 60 (1989): 157–66. See also Lawrence J. Walker, "Sex Differences in the Development of Moral Reasoning: A Critical Review," *Child Development* 55 (1984): 677–91.

61 Philip Kitcher, "Games Social Animals Play," *Philosophy and Phenomenological Research* 59 (1999): 221–8 at 225.

62 David P. Watts, "Reciprocity and Interchange in the Social Relationships of Wild Male Chimpanzees," *Behaviour* 139 (2002): 343–70.

63 Joan Silk, "The Evolution of Cooperation in Primate Groups," in Herbert Gintis, Samuel Bowles, R. Boyd, and E. Fehr, eds., *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (Cambridge: MIT Press, 2003), 17.

64 Trivers moots this friendship hypothesis in "The Evolution of Reciprocal Altruism," 48. See also Michael Argyle and Monika Henderson, "The Rules of Friendship," *Journal of Social and Personal Relationships*, 1 (1984): 211–237.

65 Philip Kitcher, "Psychological Altruism, Evolutionary Origins and Moral Rules," at 298–99, *Philosophical Studies*, 89, 2–3 (1998): 283–316.

66 Joan Silk, "Cooperation Without Counting: The Puzzle of Friendship," in Hammerstein, ed., *Genetic and Cultural Evolution of Cooperation,* 50–1.

67 Frans de Waal, *Chimpanzee Politics*, rev. ed. (Baltimore: Johns Hopkins University Press, 1998), 201.

68 Michael Tomasello, *The Cultural Origins of Human Cognition* (Cambridge: Harvard University Press, 1999), 159.

69 Bennett G. Galef, Jr., "Imitation in Animals: History, Definition and Interpretation of Data from the Psychological Laboratory," in Thomas R. Zentall and Bennett G. Galef, Jr., eds., *Social Learning: Psychological and Biological Perspectives* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1988).

70 Kathy Nagell, Kelly Olguin, and Michael Tomasello, "Processes of Social Learning in the Tool Use of Chimpanzees (*Pan troglodytes*) and Human Children (*Homo sapiens*)," *Journal of Comparative Psychology* 107 (1993): 174–86. See also Michael Tomasello, *The Cultural Origins of Human Cognition,* 29–30.

71 Andrew Meltzoff has demonstrated a similar phenomenon; see "Infant Imitation After a One-Week Delay: Long-term Memory for Novel Acts and Multiple Stimuli,"

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

Developmental Psychology, 24 (1988): 470–6. This view of imitation contradicts the "classic" view of Neal J. Miller and John Dollard, *Social Learning and Imitation* (New Haven: Yale University Press. 1941), who claim that imitation involves simply observation, instrumental reasoning, and reinforcement.

72 Ap Dijksterhuis and John A. Bargh, "The Perceptual-Behavior Expressway: Automatic Effects of Social Perception and Social Behavior," in M. Zanna, ed., *Advances in Experimental Social Psychology*, vol. 30 (New York: Academic Press, 2001), 1–40. See also John A. Bargh and Tanya L. Chartrand, "The Unbearable Automaticity of Being," *American Psychologist* 54 (1999): 462–79 at 467–8.

73 Dijksterhuis and Bargh, 15. After a survey of the evidence, they write that "in the experiments reviewed above, people did not imitate because they wanted to imitate. They imitated for no other reason than that they are designed to do so."

74 Ap Djiksterhuis, "Why We Are Social Animals: The High Road to Imitation as Social Glue," in Susan Hurley and Nick Chater, eds., *Perspectives on Imitation: From Neuroscience to Social Science*, vol. 2 (Cambridge: MIT Press, 2005), 218–9.

75 Ibid., 217.

76 For a general survey, see Andrew N. Meltzoff and Jean Decety, "What Imitation Tells Us About Social Cognition: A Rapprochement Between Developmental Psychology and Cognitive Neuroscience," *Philosophical Transactions of the Royal Society: Biological Sciences* 358 (2003): 491–500.

77 Shaun Nichols, *Sentimental Rules* (New York: Oxford University Press, 2004), 100.

78 Ibid., 119.

79 Jonathan Haidt and Fredrik Bjorklund, "Social Intuitionists Answer Six Questions about Moral Psychology," in Walter Sinnott-Armstrong, ed., *Moral Psychology*, vol. 2 (Cambridge: MIT Press, 2008), 205–6.

80 Chandra Sekhar Sripada, "Nativism and Moral Psychology," in Walter Sinnott-Armstrong, ed., *Moral Psychology*, vol. 1 (Cambridge: MIT Press, 2008), 335.

81 Ian Brockington, "Maternal Attachment and Bonding Disorders," in Susan Dowd Stone and Alexis E. Menken, eds., *Perinatal and Postpartum Mood Disorders* (Dordrecht: Springer, 2008), 24.

82 Ibid., 33–36.

83 Jesse Prinz, *The Emotional Construction of Morality* (New York: Oxford University Press, 2004), 284–5.

84 Others have noticed this. See Joshua Greene, "The Secret Joke of Kant's Soul," in Walter Sinnott-Armstrong, 76–7.

85 Terrence Deacon, *The Symbolic Species* (New York: Norton, 1998), 244.

86 Stanovich.

87 David Hume, *A Treatise of Human Nature*, 2nd ed., P. H. Nidditch (Oxford: Clarendon, 1978), 489. See also David Gauthier, "David Hume: Contractarian," in *Moral Dealing* (Ithaca: Cornell University Press, 1990), 45–76.

88 Hume, *Treatise*, 484.

89 John Stuart Mill, *Utilitarianism*, ed. George Sher (Indianapolis: Hackett, 1979), 32.

90 Benedict Anderson, *Imagined Communities* (London: Verso, 1983).

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press

91 Brandom, *Making It Explicit* (Cambridge: Harvard University Press, 1994), also Heath, *Following the Rules*, 217.

92 Joseph Heath, "The Transcendental Necessity of Morality," *Philosophy and Phenomenological Research* 67 (2003): 378–95.

93 See Christine Korsgaard, "Skepticism About Practical Reason," in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 2004).

94 For example, see Richard Joyce, *The Evolution of Morality* (Cambridge: MIT Press, 2004), 222.

95 See Eric Rasmusen, "Folk Theorems for the Observable Implications of Repeated Games," *Theory and Decision* 32 (1992): 147–64 at 161.

96 John Heritage, *Garfinkel and Ethnomethodology* (London: Polity, 1984).

97 Cass Sunstein, "Moral Heuristics," *Behavioral and Brain Sciences* 28 (2005): 531–42 writes that moral philosophy is "inadvertently and even comically replicating the early work of Kahneman and Tversky, by uncovering situations in which intuitions, normally quite sensible, turn out to misfire." See also John M. Doris and Steven Stich, "As a Matter of Fact: Empirical Perspectives on Ethics," in Frank Jackson and Michael Smith, eds., *The Oxford Handbook of Contemporary Analytic Philosophy* (Oxford: Oxford University Press, 2005), 138–46.

https://doi.org/10.1017/S0012217309990400 Published online by Cambridge University Press