Brit. J. Phil. Sci. **68** (2017), 981–1006

Evolution and Moral Realism

Kim Sterelny and Ben Fraser

ABSTRACT

We are moral apes, a difference between humans and our relatives that has received significant recent attention in the evolutionary literature. Evolutionary accounts of morality have often been recruited in support of error theory: moral language is truth-apt, but substantive moral claims are never true (or never warranted). In this article, we: (i) locate evolutionary error theory within the broader framework of the relationship between folk conceptions of a domain and our best scientific conception of that same domain; (ii) within that broader framework, argue that error theory and vindication are two ends of a continuum, and that in the light of our best science, many folk conceptual structures are neither hopelessly wrong nor fully vindicated; and (iii) argue that while there is no full vindication of morality, no seamless reduction of normative facts to natural facts, nevertheless one important strand in the evolutionary history of moral thinking does support reductive naturalism—moral facts are facts about cooperation, and the conditions and practices that support or undermine it. In making our case for (iii), we first respond to the important error theoretic argument that the appeal to moral facts is explanatorily redundant, and second, we make a positive case that true moral beliefs are a 'fuel for success', a map by which we steer, flexibly, in a variety of social interactions. The vindication, we stress, is at most partial: moral cognition is a complex mosaic, with a complex genealogy, and selection for truth-tracking is only one thread in that genealogy.

1 Realism about Scientific and Normative Thought

This article is about evolution and moral realism, and so we begin with a brief characterization of moral realism as we shall understand it, since distinguishing realism from other options is notoriously fraught. We take realism to have two aspects. One is epistemic: realists are not sceptics. A philosopher who thinks we can know nothing of atoms and their constituents is not a realist

doi:10.1093/bjps/axv060

Advance Access published on February 26, 2016

about subatomic particles. A second is metaphysical[1]: a realist about the subatomic thinks that the existence and character of subatomic particles does not depend on our opinions of the subatomic, or our concepts of the subatomic. In some sense, a realist about the subatomic thinks subatomic facts are objective, 'mind-independent'. The idea of mind independence needs to be phrased carefully when the domain in question is social. Agents' attitudes are of immense causal significance in the social world. Thus whether an action objectively harms an individual may well depend in part on that agent's view of the situation or on third-party responses. The idea, though, is that if harms are objective phenomena (as the realist understands objectivity), whether the agent is harmed is not constituted by anyone's opinions on that matter. Realism has sometimes been expressed as the thought that for a realist about a domain, it is conceptually possible for everyone, always, to be wrong in their opinions about that domain (Putnam [1975]). In debates about moral realism, this idea has been often captured by Russ Schafer-Landau's phrase: moral facts are 'stance independent' (Shafer-Landau [2003]). Moral truths are not made true by people's opinions. If it is true, as we shall be arguing, that moral facts are facts about social interactions that support stable cooperation, the moral realist must hold that cooperation-supporting institutions are morally good, independently of what anyone says, believes, or thinks. The moral realist, as we shall understand her, thinks that moral facts depend on what we are like, and how we live together and how it is that we can live together. But they are not constituted by individual or collective opinions on what those facts are. They are stance-independent. As hinted just above, we shall shortly suggest that those facts are facts about social conditions and practices that support and enhance cooperation. But as with the realist about the subatomic, the moral realist does not think that moral facts such as those are utterly inscrutable, beyond rational investigation. The rest of this article will explore the interaction between moral realism, so understood, and our best hypotheses about the evolution of moral cognition.

In the last decade, work on the evolution of moral cognition has greatly expanded (see Richerson and Boyd [2001]); Joyce [2006]; Boehm [2012]; Richerson and Henrich [2012]; Chudek *et al*. [2013]; Sterelny [2014]). What do these evolutionary hypotheses tell us about the nature of normative judgements themselves? One response, and probably the most influential, has been to take these evolutionary hypotheses to undermine the idea that there are moral facts, though there are many nuances on just how this corrosive effect should be expressed. Moral judgements are shown to be false, or probably false, or unjustified; or unjustified if taken to be about an objective domain of

[1] Thus we thus do not accept semantic characterizations of realism: the idea that to accept realism in a domain is to accept a correspondence theory of truth in that domain; see (Devitt [1984]).

moral facts (Mackie [1977]; Ruse [1986]; Joyce [2006]; Street [2006]). The sceptical idea is that an evolutionary account of the origins and stability of moral thinking displaces an account of moral thinking as responding to objective moral facts. The idea of a moral fact is shown to be redundant, playing no role in the explanation of moral belief (cf. Harman [1977]). At the same time, the argument shows that for moral thinking to play its regulative role in human social life, moralized and moralizing agents must think of moral judgements as responses to moral facts, as only this explains their power to induce agents to act in ways that often run counter to their immediate inclinations and interests.

One evolutionary analysis of religious belief is an influential model for this sceptical line of thought. On this analysis, religious commitment is adaptive, buying agents (or communities) the benefits of cooperation and social cohesion. But these benefits depend on agents' belief in the reality, power, and zeal of supernatural oversight of their actions (Wilson [2002]; Bulbulia [2004a], [2004b]). An evolutionary genealogy explains the persistence of religious belief, while showing that its adaptive benefits depend on religious commitments being taken to be truth-tracking. At the same time, it debunks it, for the analysis shows that our being prone to religious belief is not counterfactually sensitive to the existence of religious truths for those beliefs to track. We would believe in gods, whether gods were real or not. Likewise, we would have moral beliefs, whether or not there were moral facts.[2] An evolutionary genealogy of religion really does debunk religion. But religion is a poor model of moral thinking. For the effect of religious belief on social behaviour depends on agents being unaware of religion's evolved function. In contrast, while agents typically have at best a partial awareness of the evolved function of moral thinking, that awareness does not subvert its social effects.

We do not think any simple account of the origins and stability of normative thinking succeeds. We shall argue that human moral practices are a complex mosaic. Elements of that mosaic have different origins, respond to different selective forces; depend on different cognitive capacities; probably have different metanormative evaluations. After all, human moral practices include both fast, implicit, reflex-like online cognitive systems, and slow, explicit, offline systems. They involve both individual cognitive mechanisms and collective institutions (for example, communities have a stock of stories and narratives that frame their moral education). It includes both the

[2] One response to evolutionary debunking is to argue that counterfactual sensitivity of the kind characterized here is not required for reasonable belief; that is not our view (but see Enoch [2010]; Brosnan [2011]; Clarke-Doane [2012]). Another is to base moral realism on the idea that moral truths supervene on, but do not reduce to, natural facts (Jackson [1997]). For the purposes of this article, we are neutral on these options; we are exploring an alternative path, but nothing we say here undermines those other options.

internalization of individual values and the use of moral language to persuade others; we are both moralized and moralizing. The biological and cultural evolution of our moral practices very likely involved elements—norms of disgust, respect for authority, religion—that we now typically distinguish from moral thinking, properly so called (see also Machery and Mallon [2010]; Kitcher [2011]). It would be a surprise if there were a unitary explanation and assessment of all these elements. But within the moral mosaic, we shall identify one important element in the genealogy of moral thinking and argue that this strand of the genealogy of moral thinking supports reductive naturalism. Moral truths will turn out to be truths about human cooperation and the social practices that support cooperation. For moral thinking has evolved in part in response to these facts and to track these facts. So one function of moral thinking is to track a class of facts about human social environments, just as folk psychological thinking has in part evolved to track cognitive facts about human decision-making. But we do not intend or expect to deliver a full vindication of moral thinking: tracking this class of facts is only one of its functions, and the tracking is very imperfect. Our main aim is to undercut the dichotomy of vindication or debunking. Like many folk conceptual systems, moral thinking has only partially been selected as a truth-tracking system, but to the extent that its function is to track, it is neither a total failure nor full success.

The idea that connects moral thinking to the expansion of cooperation in the human lineage has two complementary aspects. First, it is important to an individual to be chosen as a partner by others; access to the profits of cooperation often depends on partner choice. Choice, in turn, is often dependent on being of good repute, and (often) the most reliable way of having a good reputation is to deserve it. It is worth being good to seem good. Recognizing and internalizing moral norms is typically individually beneficial through its payoff in reputation (Frank [1988]; Noë [2001]; Baumard *et al.* [2013]).[3] Second, human social life long ago crossed a complexity threshold, and once it did so, problems of coordination, division of labour, access to property and products and rights and responsibilities in family organization could no longer be solved on the fly, or settled on a case-by-case basis by individual interactions (Sterelny [2014]). Default patterns of interaction became wired in as social expectations and then norms, as individuals came to take decisions and make plans on the assumption that those defaults would be respected, treating them as stable backgrounds; naturally resenting unpleasant surprises when faced by deviations from these expectations. This resentment was probably recruited as one of the motives that sustain pressure

[3] This signalling or advertising function of moral response can be seen as a reason to be sceptical about truth-tracking views of moral cognition; see (Fraser [2012]; [2014]).

on would-be bullies and free riders (Sterelny [2016]). The positive benefits of successful coordination with others, and the costs of violating other's expectations, gave individuals an incentive to internalize and conform to these defaults.

These gradually emerging regularities of social interaction and cooperation were not arbitrary: they reflected (no doubt imperfectly) the circumstances in which human societies worked well, and how individuals acted effectively in these societies to mutual benefit. Given the benefits of cooperation in human social worlds, we have been selected to recognize and respond to these facts. So this adaptationist perspective on moral cognition suggests that normative thought and normative institutions are a response to selection in the hominin lineage for capacities that make stable, long-term, and spatially extended forms of cooperation and collaboration possible. On these views, there is positive feedback between moral thought and judgement and the distinctive forms of human social life. The conditions of human sociality selected for, and continue to select for, normative responses, and the emergence of norms allowed those distinctive forms of social life to stabilize and expand, further selecting for our capacities to make normative judgements.

A natural notion of moral truth falls out of the picture that moral belief evolved (in part) to recognize, respond to, promote, and expand the practices that make stable cooperation possible. For there are objective facts about the conditions and patterns of interaction that make cooperation profitable, and about those that erode those profits. For example, Elinor Ostrom has identified general characteristics that make collective action problems more tractable (Ostrom [1998]). There are also objective facts about the practices and norms that would promote stable cooperation within the group. Evolutionary game theory is helpful here, since its analysis often shows that distinct equilibria—different stabilized patterns in behaviour that become customs and norms—differ in their capacity to deliver cooperation profits. No doubt there is no single set of optimal norms; the best normative packages for a group will depend on its size, heterogeneity, and way of life. No doubt there are trade-offs between the size of the cooperation profit and its distribution. But despite these complications, a natural notion of moral truth emerges from the idea that normative thought has evolved to mediate stable cooperation. The ideal norms are robust decision heuristics, in that they satisfice over a wide range of agent choice points, typically providing the agent with a decent outcome, in part by giving others incentives to continue to treat the agent as a social partner in good standing. The moral truths specify maxims that are members of near-optimal normative packages—sets of norms that if adopted, would help generate high levels of appropriately distributed, and hence stable, cooperation profits.

On this view of moral thinking, as with other neo-conventionalist accounts, moral thinking emerges as a version of prudence (in this respect, our views are

similar to those of (Gauthier [1987])). In general, agents have an individual stake in supporting effective yet fair cooperative practices. We might prefer unfair solutions if we were to be part of the elite, but fairness typically satisfices. A fair social world might not be our first choice, but it is certainly not bottom of the list. We are moral only because for most of us most of the time it was and is in our interests to be moral. But our evolved dispositions make us genuinely moral. Moral response is not voluntary, not conditional on individual decision or calculation at particular choice points. To borrow a term from Daniel Dennett, our commitment to moral policies is ballistic, rather than being re-evaluated on a moment-by-moment, case-by-case basis; we are at least somewhat moved by our moral opinions, even when we would rather not be (Dennett [1995]). We do not decide on a case-by-case basis to feel moral emotions or to make moral judgements. Sometimes, then, thinking and acting morally will not be in an agent's interests. But because such cases are atypical, we have been selected to genuinely endorse moral views and to make them part of our motivation structures, even though it would sometimes be in our interests to ignore them. As with any form of naturalism, on this view of moral thought, the motivational force of any moral claim is extrinsic to its content. Moral facts exist independently of any specific agent's recognition of those facts. Their power to motivate us is contingent, but derives from developmentally and evolutionarily deep and relatively inflexible features of typical human personalities. Objective moral authority, if it is anything, is ballistic commitment; it is part of the point of this article that this is neither explaining what moral authority is, nor explaining the illusion of moral authority.

In this article, then, we have three targets: First, we aim to locate evolutionary debunking and reductive naturalism within the broader context of the relationship between science and the folk frameworks for thinking about the world. Second, we shall suggest that these other cases undermine the dichotomy between reduction and debunking. Folk conceptual frameworks can be imperfect, yet still latch on to and partially describe important phenomena in our environments, and guide action with respect to those phenomena. Third, we argue that the mosaic character and the complex genealogy of moral thinking and practice are important.

In particular, we shall suggest that while moral thought and judgement in part evolved to facilitate mutually profitable social interaction by tracking and responding to roadblocks that limit cooperative profits, this is only one factor in the matrix of selection through which human moralizing emerged. In Lewis–Skyrms signalling systems, signalling emerges when one agent, 'the receiver', can act, but lacks information about the environment to which only a second agent, 'the sender', has access. If the receiver acts to their mutual benefit, signalling emerges. These signals both track variable states of the environment and guide adaptive response, and this is the core from

which indicative, truth-apt language was built (Lewis [1969]; Skyrms [2010]). But signalling systems can emerge as pure coordination devices, when agents have an interest in mutually adjusted interaction, as in dance and many games–but where the benefits of coordinated interaction do not depend on the coordinating signals' match to some independent, variable feature of the world. In these cases, the signals do not have any world-tracking function (Godfrey-Smith [2012]). Linguistic conventions about (for example) word use are coordinating devices like this, and we shall argue that moral norms play this pure coordinating role too—one that does not depend on them tracking independent features of the social world. They may well have other roles too aspects of sexual display, or as devices for both marking and deepening ingroup–outgroup distinctions.

This complex genealogy is relevant for two reasons. First, to the extent that these non-tracking functions are important, we would expect genetic and cultural selection to be less effective filters: less effective in predisposing us to norms that actually promote profitable, stable cooperation. Second, to the extent that these non-tracking functions are important, evolutionary considerations speak less persuasively in favour of a cognitivist view of moral thought and talk. If moral thinking evolved as a tracking device, selected to track and respond to cooperation pitfalls, then the apparently truth-apt character of moral thought and talk would reflect its functional role. The less it evolved as a tracking device, the less its apparent form reflects role, and the more plausible non-cognitivist options become.

The road ahead is as follows: The next two sections reflect on reductive naturalist hypotheses in general—when should we discard folk frameworks and when should we regard them as largely vindicated by our best science? We conclude with an intermediate case: folk astronomy. We think folk astronomy is important, because it supports adaptive action quite flexibly, despite astronomical belief being embedded in a seriously mistaken set of general beliefs. In our view, folk astronomy is a good model for normative belief. In Section 4, we respond to the argument that moral facts are epiphenomenal; that while moral opinions are causally important, moral facts explain nothing. In Section 5, we make a positive case for the explanatory importance of true moral beliefs: we show that in some important ways, they are maps by which we steer. We then summarize and conclude.

2 The Folk and Science

One of the great projects of contemporary philosophy is to explore and identify the relations between two apparently different ways of thinking about humans and their place in the world. We develop one set of cognitive tools from our socialization as members of our communities: we develop folk

understandings of the physical world, the biological world, human agency, and so on. This is Sellars's 'manifest image', though we should of course speak of manifest images, as there is no single folk framework; it has varied across time and culture. The other conceptualization has developed within the natural and social sciences over the last few centuries: the 'scientific image' of humans and of the world in which they act. The two conceptualizations are not obviously compatible. For example, the view that humans are self-aware, rational decision-makers is not obviously compatible with the view that we are modified great apes. If the views are not compatible, how should we respond?

One major move in this philosophical space is reductive naturalism. The key strategy is to co-opt an idea developed in understanding the relationship between sciences, and use it to understand the relationship between folk thought and scientific thought. Within science itself, reduction is the claim that the facts in one domain—a reduced domain—are less explanatorily fundamental than the facts in a reducing domain. Facts about inheritance patterns between parents and offspring in sexually reproducing population—the facts systematized and predicted in classical genetics—are less fundamental than facts about the sequences of DNA base pairs in the haploid gametes transmitted across generations, the gametes that fuse to form a new individual. Facts about the stability of a species over time are less fundamental than the facts about the flow of genes in that species' gene pool, and their constrained flow outside that gene pool. The reducing facts explain the reduced facts, but not vice versa. The most plausible cases within the sciences (perhaps the only plausible cases within the sciences) are relations of composition. The character, distribution, and interaction of the parts of a system explain the behaviour of the system as a whole.

Reductive naturalists extend this idea to folk kinds. Famously, at the dawn of this project, Jack Smart suggested that facts about conscious experiences of pleasure and pain were explained by, and reduced to, facts about human neurophysical organization and activity (Smart [1959]). Reductive naturalists point out that the reductive relationship between domains—even when it involves kinds recognized by the folk—is not in itself a piece of folk wisdom. For these reductive relations between domains are discovered empirically, they are not *a priori* or conceptual truths. Nor are they obvious truths of unaided observation. One of the stock examples is the identification of water as H_2O; now widely known as a chemical factoid, and an example of compositional reduction, but once a major discovery of scientific chemistry. This is no surprise on a system–component model of reduction. The folk will often be acquainted with, and have reliable information about, complex macroscopic systems—like organisms or agents—but be without systematic access to information about their internal components and their organization. So naturally humans can develop a concept of water, and know lots of truths about

water, without knowing that it is nothing but a configuration of oxygen and hydrogen atoms.

The folk can have epistemic access to a system without thereby having epistemic access to its components and organization. A second model, exemplified by the notorious identity of the morning and the evening star, depends just on this idea of distinct routes of epistemic access. The thought here is that we can have epistemic access to the same individual or kind through two different routes, and we can form a variety of true judgements about a kind or its instances without realizing that there is only a single kind in play. Early versions of materialist theories of mind—early forms of functionalism and Smart's 'topic neutral' analysis of mental concepts—had to explain why the identity of mind and brain (if they were indeed identical) was not obvious to us all. Their response was initially ambiguous between a two-routes model of epistemic access and a system–component model. But later versions of functionalism—the functional decomposition models of Dennett, Lycan, and Stich—are clearly system–component models (Dennett [1978]; Lycan [1990]; Nichols and Stich [2004]). The point, though, is that any form of reductive naturalism targeted at folk kinds needs some account of how a reductive identity can be true, without its being known to be true, despite the fact that folk agents have plenty of information about the reduced domain, and sometimes even the reducing domain.

Norms are part of the manifest image. We think of actions as cruel or kind, generous or stingy, required or forbidden. We think of some people as admirable, and others as arseholes. To put it floridly, typical humans take themselves to live in a normative world, not just a physical world. How does this aspect of the manifest image relate to the scientific image? There is a version of the reductive naturalist project, known as 'Cornell realism' (Boyd [1988]; Brink [1989]), that extends that project to normative phenomena. Norms turn out to be natural facts.

The Cornell realists take 'water = H_20' as their paradigm for thinking about the relationship between natural and normative facts, because this paradigm blocks the open question argument. It shows that a reductive identity can be true without its truth being apparent to any cognitively and linguistically competent member of a community. But in other respects, it is a misleading model. There is no compositional relationship between normative kinds and any plausible set of base properties. The reduction base is a set of facts about agents, their interests, the social systems of which they are a part, and the deep history of those social systems. So, it is important for the reductive naturalist project that there be another, non-compositional kind of reduction relation between the normative and the natural.

Dan Dennett's picture of the intentional stance offers a better model of the relationship between the normative and the natural. The cognitive and neural

organization of an agent sharply constrains the belief–desire profiles we can attribute to that agent (Dennett [1991]). So there is a very important relationship between belief–desire psychology, cognitive psychology and cognitive neuroscience. But while these constraints are important, they do not uniquely specify an intentional profile. Dennett argues that no agent's actual behavioural dispositions will perfectly match any intentional profile; such profiles always idealize behavioural patterns to some extent. Profiles can be distinct, but equally legitimate, because they make different trade-offs between simplicity and accuracy. Moreover, while the cognitive organization of an agent explains their behavioural dispositions and hence their intentional profile, specific beliefs and desires do not routinely map on to specific elements of an agent's cognitive organization (see especially Dennett [1991]).

The relationship between intentional and cognitive psychology is a better model for the realist of the relationship between natural facts about human social life and normative claims, because it is not a system–component view of the relationship between domains (beliefs, for example, are not composed of specific neurocognitive structures). This picture does not commit us to the view that there is a unique set of moral truths fixed by the reduction base, nor does it commit us to the view that there is an element-by-element reduction of normative predicates to natural predicates. That is important. On the hypothesis we have been considering, the reduction base is a set of facts about human communities (including ancient ones): facts about profitable forms of cooperation, about social arrangements and cognitive dispositions positively and negatively relevant to the stable exploitation of those opportunities. These complex social environments selected for human recognition of, and response to, maxims of social interaction, which in general improved human access to these benefits. But it is most unlikely that there is an element-by-element mapping between the norms in adaptive packages, and opportunities and barriers to cooperation. Norms are typically relevant to many action choices in many contexts. So in thinking about the relation between norms and facts, we will take Dennett's model rather than Putnam's as our guide.

3 Reduction, Vindication, and Error

The reductive project, when carried through successfully, is intended to vindicate the folk conception of the world. A theory of free will, for example, might identify free action with informationally sensitive decision-making guided by stable motivation, and show that on important occasions humans make decisions of that kind. Such a theory would vindicate the idea that human agents sometimes act freely. Contrast that with a sceptical theory

that emphasized our ignorance of our own motivational structure, our cognitive biases, and the sensitivity of action and judgement to clearly irrelevant contextual factors. A theory of this kind would be best seen as explaining the illusion of free will. But this free will example raises a methodological challenge to the project of naturalistic mapping. How should we distinguish genuine from ersatz mapping? Thus it's often claimed that the folk are committed to the idea that there is real free will, real autonomy, not merely the (approximate, fallible) capacity to act with appropriate informational sensitivity on the basis of a stable, reflectively assessed preference order.

The ersatz problem makes it natural to link the project of naturalistic mapping to one of philosophical anthropology. The idea is to take a domain of folk opinion (in our case, normative thought and talk) and attempt to systematize that opinion. This project proceeds by a mix of methods. Ideally, a systematization of (say) the folk concept of consciousness will capture both the folk's intuitive judgement of specific cases—I am conscious right now as I read this article—and the folk's general principles about consciousness. For example, it will capture the idea that consciousness is psychological state, but not one agents are in all the time; whether you are conscious of a particular event is relevant to whether you enjoy it; adult humans engaged in ordinary mundane activity are conscious; rocks and corpses are not conscious; and so on. The 'Canberra plan' is a particularly well-developed and theoretically well-motivated version of philosophical anthropology (Jackson [1997]). The Canberra plan is alert to the fact that we should not expect the folk to be completely unanimous: we should expect there to be marginal or debatable examples of folk maxims (in this case, perhaps whether consciousness comes in degrees), and we should expect some failure of fit between judgements of particular cases and the systematization of folk principles.

Moreover, as is standardly recognized on this kind of approach, the maxims need not all be of equal importance. Perhaps in the case of consciousness, maxims about the relationship between conscious experience and affective valence are more central than those about non-inferential knowledge of conscious states. It is also true that folk maxims, especially the general principles, are rarely explicit features of folk frameworks. The project makes implicit commitments explicit, typically by reflection on intuitive judgements about particular cases—a procedure that leaves plenty of room for uncertainties. So there are problems and complications, but once we have identified the core folk commitments about (say) consciousness, we have, in effect, constructed an implicit definition of consciousness. For all the clear, core, uncontroversial features of the folk's view of what it is to be conscious, it is that unique state X that satisfies the following conditions: X is mental state; awake, normally

acting adult humans are in *X*; rocks are never in *X*; and so on. Once we have done that, we have constructed a potential bridge between putative folk kinds and our best science. For we can then ask, from the perspective of our best science, whether there is any unique kind that satisfies the conditions specified in the implicit definition. Philosophical anthropology might show that stable, self-reflective rational decision-making is necessary and sufficient for free action. It would then be the task of cognitive and social psychology to determine whether human decision-making regularly (or ever) satisfied these conditions and, in particular, whether it does so in those cases regarded as paradigmatic of free action. If not, we should be error theorists about free choice, as we actually are about witches.

Thus the Canberra plan is alert to the possibility of error. The folk might be irreparably wrong in some central aspect of their thinking about the world, and some folk frameworks have rightly been discarded, but one of the strengths of the Canberra plan is that it very naturally recognizes the fact that there are many cases intermediate between vindication and error theory. There may be a unique state that satisfies some but not all of the clauses in an implicit definition of free will or conscious thought; there might be a state that satisfies all or most of the conditions, so that some human actions are free, but not those cases taken to be paradigms of free action. Suppose, for example, that agents often make very good decisions when they make fast, on-the-spot judgements in situations in which they are very experienced, showing just the right sensitivities to subtle differences in circumstances. But when they attempt to make good decisions through explicit, slow, careful self-conscious reasoning, they are especially prone to framing effects and irrelevant contextual cues. The idea that we make free choices would then be neither vindicated nor debunked. Likewise, suppose that once we regimented folk opinion about moral norms, it turned out that no natural social phenomenon matched the folk conception of the moral. Perhaps (for example) it turns out that a neo-Kantian maxim—a moral norm gives agents an overriding reason to conform to that norm, irrespective of their preferences—is central to the folk conception of a moral norm. It would follow, on the Canberra plan, that there are no moral norms or moral facts. But it would not follow that there are no quasi-moral or moral-like norms, that there is nothing in the social world that somewhat approximate to the folk conception of a moral fact.

So vindication and partial vindication can come from being able to map folk frameworks on to scientific frameworks. However, despite its capacity to recognize intermediate cases, the Canberra plan seems to over-count failed folk frameworks. So we need an account of vindication and partial vindication that incorporates ideas from the Canberra plan, but which goes beyond it. Consider thought and talk about the stars and planets in ancient world

(the thought systematized and quantified in the Ptolemaic astronomy).[4] A systematization of Mediterranean astronomical thought of AD 200 might suggest that we should be error-theorist about pre-Galilean astronomy. Almost all of the general beliefs were mistaken, as were some of the particular identifications (the moon and sun are not planets; the earth is). Yet that does not seem right, for agents in the ancient world were able to use astronomical information adaptively in navigation and to tell the daily and seasonal time. Of course, there is wriggle-room. For example, the Canberra planer can insist that the maxims with the heaviest weight are ones like, 'You can only see the stars at night', or, 'the stars do not seem to move in their relative positions but the planets do', or, 'Mars looks reddish when it is brightest'. But this does seem to shift away from the idea that the agents in question implicitly had a coherent conception of the night sky, a conception that we can systematize and then attempt to map on to our best scientific conception. For it does not seem plausible that the agents themselves would have regarded these banal maxims of perceptual observation as their most central astronomical beliefs, especially once astrology took hold of both the lay and the educated mind.

We think that this example shows the importance of know-how or skill, and suggests that skill is not just a special case of propositional knowledge. Sky-watchers of the ancient world had a complex of explicit beliefs about what they could see, but they also had a complex of discriminative capacities. They could identify and re-identify specific celestial objects and configurations, and those discriminative capacities both fed into descriptive beliefs and supported adaptive action, for example, direction-finding. Folk cognitive frameworks, on this view, are not just systems of propositional representation, and these frameworks can enable agents to register features of their environment and guide response to them in ways that partially screen-off mistaken belief, sometimes even when those mistakes are quite fundamental. The folk can sometimes respond in quite nuanced ways without that response being routed through a conceptualization of the phenomenon in question (presumably non-human animals typically manage their environment this way). Moreover, some of the beliefs depend quite directly on the discriminatory capacities and these can guide adaptive action, for example, knowing the tides are higher when the moon is full. Folk frameworks can be responsive to real phenomena, and guide action appropriate to those phenomena, without accurate conceptualizations of those phenomena. The ancient world registered and responded to features of their celestial environment, and this guided navigation, calendar construction, and time-keeping. Our point is not just that

[4] Thus this example is not strictly speaking a folk framework, since it includes elements that are produced by cultural elites—like calendars and almanacs—which are then absorbed into the general practice of the community. We do not think this complication affects the example, as folk frameworks often incorporate new elements.

flawed folk frameworks can be practically useful; they are useful because they are not just wrong, though they are not just right either. We think something similar is true of moral response, especially automatic, reactive moral responses. These depend on implicit generalization from exemplars, rather than on explicit principles of moral reasoning, but this is still a mode of representing the social environment. Moral cognition is partially know-how; it is not just a structure of propositions (Stich [1993]; Churchland [1996]; Sterelny [2010]).

Discriminative capacities, and the banal but true beliefs that they support, help distinguish ancient astronomy from other apparently mixed cases. For example, taboos often support adaptive behaviour (Harris [1985]), but in rigid and limited ways. In certain Amazonian tribes, fish-eating fish are a taboo food for pregnant women (Begossi *et al*. [2004]). As it turns out, these fish contain high concentrations of toxins in virtue of being near the top of their food chain. So the tribesfolk are acting adaptively in identifying the fish as having this apparently spooky property and so avoiding it, but one might not think this much of a vindication. Suppose, though, that in addition to avoiding these fish, these agents have a way of identifying the toxin wherever it is found (suppose it to have a distinctive colour when baked), and always avoid it. The practice would still be embedded, as with ancient astronomy, in a deeply mistaken theoretical framework, but with the support of these discriminating capacities, identifying taboo substances would support adaptive action quite flexibility (it would be a fuel for success). Flexibility of adaptive response is one dimension along which folk beliefs and practices can vary. The taboo case would then be like some of the more successful elements of ancient and folk medicine (for an analogous case, see Henrich and Henrich [2010]). For some diseases and injuries have long been identified, and to some extent effectively treated, despite these practices being embedded in very mistaken theory. Malaria became such a case. By the seventeenth and eighteenth century, European physicians were aware of the connection between malaria and exposure to wetlands, and the use of quinine was becoming standard. But they had no clue about the aetiology of the disease or the cure: 'malaria' derives from 'bad air', and it was thought that vapours rising from swamps caused the disease. Equally, quinine had been introduced as a lucky guess; South America Indians used it to reduce shivering when they were very cold (Rocco [2000]). So malaria is another mixed case, but not all cases are mixed.[5]

[5] One of the readers asks whether we are compelled to treat phlogiston theory as a mixed case, given that the defenders of that view of combustion were able to manipulate combustion in quite sophisticated ways. Very likely, we are, but we do not regard this as an embarrassment: it has been convincingly argued that a closer look at the history of chemistry shows that the phlogiston theory of combustion is indeed a mixed case (see Kitcher [1993]).

As we noted in the introduction, there is a line of thought that suggests that religious belief prompts adaptive behaviour and in particular, prosocial, cooperative behaviour. We are not convinced. But even if religious belief is adaptive, it does not leverage adaptive behaviour in the ways ancient astronomical lore did. The pre-modern conception of the solar system leveraged adaptive action with respect to navigation, time-keeping, and season-tracking only because it counterfactually tracked some structural and dynamic features of the solar system quite precisely. Had, say, the relationship between day length and seasonality been different, first, that would have been relevant to the success and failure of actions, and second, the Pre-moderns would have noticed. In contrast, a striking fact about selective models of religion is that the details of the religious system endorsed by the agents do not matter: at most, all that matters is that believers think there are powerful, though hidden, supernatural police. Procedural details—does a doorman guard the gates to heaven, or is one's heart weighed against a feather?—don't matter. Similarly, to recycle a clichéd example, seventeenth-century witch theory was a folk framework that deserved elimination. Even if those persecuted were an identifiable sub-group (friendless, isolated, socially deviant), rather than an *ad hoc* collection of the unlucky, discrimination did not leverage adaptive behaviour, even by the lights of the witch-burners. It did not prevent crop failures or other misfortunes. Perhaps (though we greatly doubt it), witch-thought paid its way by building a stronger sense of community amongst the survivors. Even were that so, the details of the ideology would not matter: persecuting outsiders for bargaining with the devil, for having tainted blood, or for being alien zombies would all work as well in building community. There are no dependencies between details of the framework, features of the environment, and adaptive choices. So nothing in the world remotely corresponds to the witch-identifying maxims, nor did witch representations leverage adaptive behaviour.

So vindication is possible, and so is elimination. But we think that the most important upshot of this discussion is that it is a mistake to frame the discussion of folk ontologies as a choice between reduction and elimination. In many cases, that framework is misleading. For one thing, folk frameworks are part of the folk's world; they are not just tools for navigating that world. Folk psychology is an active ingredient in human developmental environments (Mameli [2001]; Ross [2006]; Zawidzki [2013]); likewise, the norms that are expressed, taught, and enforced are important aspects of human social environments. Even when we focus on their role as maps, many cases—perhaps most cases—will involve some mix of vindication and rejection, some of mere causally grounded responses to phenomena in the world, of partially correct conceptualization and description of those phenomena, and some capacity to support effective action through tracking and conceptualization. In particular, the ancient astronomy example shows that folk conceptual systems can

systematize responses to phenomena in the world in ways that leverage adaptive behaviour, even though those conceptual systems misdescribe their targets in genuinely important ways. Despite the errors in pre-modern astronomy, pre-modern astronomical beliefs leveraged adaptive actions from those own agents' own perspective regularly, systematically, and non-accidentally. Pre-modern beliefs about witches did not, since burning witches did not stop crop failure, plague, and other local disasters, or even expel the devil. From the perspective of the witch-finder's own ends, witch killing was ineffectual. So witch lore did not give agents theoretical leverage over the nature of the world, nor did it give them practical leverage in making things happen. Ancient astronomy gave a little of the first, and quite a lot of the second. Ancient astronomy, then, is a mixed case.

4 Moral Facts and Moral Opinions

Prima facie we would expect folk moral theory to be at best a mixed case, too. We noted in the introduction its mixed genealogy. While moral thinking evolved to track the social environment, it did not evolve only as a tracking device. Selection did not favour believing all and only the moral truths. Moreover, hidden-hand mechanisms are far from perfect in producing optimal adaptations to heterogeneous and fast-changing environments (Sterelny [2007]). The wide variation in moral opinion seems to confirm this pessimistic expectation, showing that if moral thinking tracks moral facts, it cannot be doing so very efficiently. Perhaps some variation is a sensible adjustment to different local circumstances, but much cannot be. One source of error is that, as with ancient astronomy, in many cultures moral thinking keeps bad company, being entwined with bizarre religious misconceptions, local origin myths, dubious politics, and crackpot notions of purity and health. In addition, there is often at least some self-serving influence of elites on local moral opinion. So no adaptationist, truth-tracking conception of the evolution of moral thinking will deliver a full, clean vindication of diverse moral opinion. Indeed, we expect the moral case to be intermediate in a variety of respects: First, our moral practices are a mosaic; some elements may turn out to be vindicated, others revised, others discarded. Second, as we have noted, moral judgements function to signal, to bond, and to shape, not just to track; vindication is only in question with respect to tracking. Third, as we shall now explain, tracking is only partially successful; moreover, its success may well have varied across time and circumstance. For example, it is possible that in the intimate, informationally transparent, and relatively egalitarian social life of Pleistocene foragers, normative thinking tracked cooperation-supporting customs, and leveraged adaptive action more

effectively than it did in the much more hierarchical, elite-controlled social world of the first states.

To even partially vindicate folk moral theory, the evolutionary realist must meet two challenges. First, the debunkers have argued that the appeal to moral facts or moral truth is redundant: we can explain human moral thought, and the influence of thought on action, without appeal to moral facts. Second, the evolutionary realist needs to develop a positive case, analogous to the one noted for ancient astronomy. In thinking about astronomy, we saw that despite theoretical misconceptions, many folk astronomical beliefs were true (even though they were pretty mundane observational beliefs) and that the cognitive network of astronomical beliefs, and the perceptual capacities that supported them, powered adaptive action quite flexibly and over a range of contexts. Folk representation of their celestial environment was, to some extent, a 'fuel for success': it advanced the interests of the agents who employed it (note, we are here talking about utility and not fitness, two things which may go together but can come apart). Can we show the same about folk moral thinking?

Moral realists do not deny the existence of individual or collective error. So moral facts, as the realist understands moral facts, do not play a privileged role in the genesis of every moral opinion. Rather, on the hypothesis that moral truths are truths grounded in facts about cooperation, the project is to show that in favourable cases there is a reliable causal connection between moral opinion and these facts. The critical constraint is to provide a natural history of moral opinion formation that identifies distinct causal pathways in the immediate psychological history of individuals, or in the social context of social learning, or both. The natural history of true and partially true moral opinions must be systematically different from false ones, and that difference must involve moral facts playing a regulative role. There are, of course, some common features in the genesis of moral opinion. All moral learning involves an interaction between our systems of social emotion, individual trial-and-error learning (as children explore in, and negotiate, their social space), and the moral opinions of their community. These community opinions are expressed tacitly in their actions and interactions with one another; less tacitly in their customs and institutions; explicitly in their normative vocabulary, explicit moral maxims, and narrative life (for a more detailed exposition of this view of moral learning, see Sterelny [2010], [2012]). However, though all social norms are acquired by some form of social learning, not all learning pathways are equal. The evolutionary moral realist can make it plausible that there is indeed a systematic difference between the history of error and of truth. While every moral belief is the result of social learning, not all social pathways to belief are equal.

In particular, we suggest that there are three ways that agents become aware, with some reliability, of the opportunities and challenges of human cooperation, and come to endorse norms that improve access to the profits of cooperation: (i) learning guided by prosocial emotions; (ii) vicarious trial-and-error learning in heterogeneous environments; and (iii) cultural group selection. We begin with prosocial emotions.

Jesse Prinz and Shaun Nichols argue that while norms are learned socially, some norms are especially salient. There is a particular learning route that goes via our recognition of emotional response: cases where our acts affect other agents about whom we care, and we notice both their emotional responses to our actions, and our emotional responses to their responses. Thus generosity to others is readily reinforced through a loop in which their positive response induces your own positive response through emotional contagion; a response that you yourself notice. This does not guarantee the acquisition of norms of sharing (nor norms of harm avoidance, in the negative case). But it does make the phenomena that fall under those norms salient (Nichols [2004]; Prinz [2007]). Salience is no guarantee of truth. But our species has had a long history of biological and cultural selection in favour of cooperation-supporting emotional responses. Patterns of behaviour we find emotionally repugnant are likely to be instances of behaviours that would be forbidden by socially efficient norms. Patterns of behaviour we find appealing are usually instances of behaviours that would be endorsed by socially efficient norms.

Second, many contemporary societies are normatively heterogeneous, composed of cross-cutting groups with competing norms and agendas. In these heterogeneous contexts, agents have some ability to treat each other as natural experiments. In interacting with others who embrace different normative packages, we have some opportunity to see how their lives go. For example, do they live in networks of support and mutual aid, are they regularly exploited by freeloading neighbours, or are their lives blighted by mortal feuds and the enmity of former friends? While a full-blown evolutionary perspective on the origins and stabilization of moral cognition is not part of folk wisdom anywhere, the idea that norms have a social role that promotes fair interaction may well be, and in these mixed learning environments, that awareness may play some role in the norms agents accept and internalize. After all, moral education quite often proceeds by noting the effects of norms, and hence norm violations, on cooperative lives ('What if everyone did that dear?').

In this respect, moral norms are very different from the religious norms we discussed in Section 1: the social role of moral norms can be transparent to end-users without that knowledge eroding their role. While evolutionary models of the emergence of norms do not presuppose that agents understand the role norms play in their lives, they do not presuppose that they have no

insight into this role. Indeed, because normative facts are mundane facts, ordinary agents have access to many of them, and so folk reflection is not futile. On this evolutionary naturalist picture, there is nothing mysterious about moral epistemology. That would be different if, say, the truth-makers for normative claims were historical facts about the Pleistocene. Moral knowledge is not knowledge of mysterious or inaccessible facts. That is important: as we noted in Section 1, a moral realist had better have an account of how moral knowledge is possible.

Third, in the past, communities were smaller and more internally uniform. Some of these communities did well; others, less well. Arguably, one causally relevant factor was the extent to which their normative lives stabilized and enhanced local cooperation. Cultural group selection will favour systems of moral norms that are relatively efficient means to the ends of social peace, regulation of conflict, and the restraint of selfish or destructive impulses (Boyd and Richerson [1990]; Bowles and Gintis [2011]; Chudek *et al.* [2013]). Evolutionary naturalists should see the evolution of norms as an ongoing process of gene–culture coevolution. But this is independently plausible: human cognitive and cultural evolution did not stop in mid-Pleistocene Africa.

These moral truth-tracking mechanisms can be overridden by other processes, and even when they guide norm acquisition, they are by no means guaranteed to guide agents to true norms, to one of the optimal packages. But given a social and physical environment, and a set of interacting agents with their opinions and motives, there will be facts about whether their current norms are efficient means to stable and profitable cooperation. And to the extent that norms that do support cooperative interactions become established in a culture, it is typically not just by lucky accident. There is some tendency for better norms to be found, though this process is noisy, imperfect, and dependent on deep evolutionary histories, not just intelligent individual learning.

If this is right, actual systems of moral opinion will be a mixed bag. The naturalist project is to show that the elements in this bag tend to have rather different cultural histories, and depend on different social learning processes. Some will be unfortunate historical legacies (lingering prejudices of various kinds). Some will be levers for exploitation and injustice that exist because of imbalances of power and wealth. Some will indeed be the result of selective filtering, but not for tracking and responding to levers of cooperation. But some actual maxims will be true and their truth will have played an important role in their becoming widely endorsed. Had the natural bases of cooperation been different, moral beliefs would have been different. For example, it is surely likely that the maxim, 'do not be cruel', or the maxim, 'do not inflict severe pain for fun', will be part of most packages of norms that promote efficient and stable cooperation. Cruelty is no longer offered openly as public

entertainment. That is a change since the days of public execution, and it is a change propelled, in part, by the acceptance of an anti-cruelty maxim, and that maxim has been accepted because it is true. The truth of the maxim is not causally idle: it is relevant to its presence, persistence, and learnability. Suppose that humans had a different motivational psychology. For example, suppose that a cooperative world was possible only with wide social support for cruel punishment. Then (we conjecture), our normative response to cruelty would be different. To the extent that our moral opinions causally depend on the social and psychological bases of stable cooperation, they do so in a counterfactually sensitive way.[6]

Even if we accept the view that there are importantly different routes through which moral norms come to be endorsed and internalized, even in favourable cases, there is a striking difference between moral response and cases where automatic response is guided by, and developed through, a reflective understanding on the phenomenon being assessed. Philip Kitcher, for example, has pointed to a sharp contrast between our moral responses and, say, a physicist reading a bubble chamber photo (Kitcher [2011]). When an expert scans bubble chamber photographs, while the assessment is automatic, the practitioner knows the key elements of the vindicating history; the trained eye is supported by theoretical reflection. This is not true of intuitive moral response. But even in science, reflective vindication is an achievement of maturity. It is not in place at the beginning of the process. Consider, for example, biological classification before mature evolutionary biology. Linnaeus built on existing practice, but from his work, biological systematics flourished, with organisms being identified and described, and sorted in to species, genera, and family. By our current lights, these practices were quite reliable. But the practitioners lacked a vindicating theory of their practice; they lacked, for example, a vindicating theory of homologies and how they were to be distinguished from other forms of similarity, though their actual methods were quite reliable. The history of systematics shows that it is possible to respond to and track a phenomenon (in this case, the tree of life) without a good account either of the nature of the underlying phenomenon or of why

[6] One reader suggested that cruelty-permitting or encouraging maxims might be part of some optimal packages, especially when one remembers the context sensitivity of optimal packages. This question raises delicate issues. It is certainly true that agents can have reason to endorse and internalize cruel packages; as we note in the final section, norms play a coordinating role as well as a tracking role, and agents have reason to coordinate with the norms of their community, even the ill-chosen ones. But a complication is that 'cruel' is already a normatively loaded term. Suppose we strip that out and replace it by a more descriptive locution, say, 'taking pleasure in the pain of others'. Then, as we note, it is not so hard to imagine a world in which such a norm might be part of an optimal package, if it is part of the normative machinery that recruits the punishment of freeloaders and bullies. It might be normatively appropriate to enjoy the suffering of those that deserve to suffer. We are no theologians, but we are told than on some versions of Christianity, one of the pleasures of heaven was to spectate the sufferings of the dammed.

the perceptual proxies are in fact good signals of that phenomenon. It is true and important that those who debated, and continue to debate, moral choice often have no good account to offer of the nature of appropriate moral maxims, nor of the evidence that supports one view over another. But the same was true of scientific pioneers. Reflective understanding is an achievement of maturity.

5 Is Moral Knowledge a Fuel for Success?

Moral language has the form of a fact-stating discourse: 'Stalin was cruel' and 'paedophiles deserve to be locked away' have the form of ordinary indicative sentences. That does not show much. We do not need a robust, correspondence notion of truth to explain the logical or inferential roles of truth; for that, deflationary theories suffice (Horwich [1998]). Simon Blackburn and Philip Kitcher have developed theories of moral language and cognition that treat moral language as indicative, but without any serious commitment to moral facts. We need a substantive notion of truth when the representational properties of language and thought help explain success, when that success is flexible, and when the representational capacities support adaptive action across a range of projects. We need a robust, explanatory notion of objective fit between mind and world to explain systematic success of thought-guided action; when beliefs that accurately represent the world are a fuel for success (Godfrey-Smith [1996]; Sterelny [2003]). At the end of Section 3 we argued that agents who used the framework of ancient astronomy in representing their celestial environment thereby built a mental map that was to some significant degree a fuel for success, despite the theoretical flaws of the framework and despite its incorporation into magical modes of thinking. Are moral beliefs likewise fuels for success when, and in virtue of being, true, and despite the fact that they are often enmeshed with superstition and prejudice?

We see a case for a partially positive answer. But the mixed genealogy of moral thinking is also important. Moral norms often play the dual role of coordinating devices and cooperation amplifiers, promoting choices that give other agents incentives to cooperate in turn. These roles can conflict. For once default forms of action become established in a community, agents have incentives to conform to them, even if they eliminate or erode cooperation (Boyd and Richerson [1992]). Agents have incentives to match their normative beliefs to those of their community, whether those beliefs are true or not. As we remarked earlier, moral norms are not just reflections of our social environment, they are features of that environment. Adherence to local norms is part of the process that establishes common knowledge: sets of background expectations about others and how they will behave, and expectations on which agents rely in planning and coordination. If local defaults rule out

social interaction between the sexes, violating those expectations will at best cause coordination failure and social uncertainty; the agent who does not act as if females were potential sources of pollution is weird, unpredictable. Typically, there are even stronger incentives to conform, because the normative views of an agent are themselves the subject of normative assessment. Part of being moral is having the right moral beliefs. It is not enough to avoid paedophilia; one must also think that paedophilia is wrong. It is not enough to avoid talking to women; one should think that talking to women is wrong. In contrast, folk astronomy was not especially a tool for coordination and social interaction, and except when they became enmeshed in religion and magic, folk astronomical beliefs were not socially marked. So, there was no special pressure to conform to others' errors.

Moral thinking is not a domain in which, all else being equal, the true belief is automatically rewarded. Even so, truth—identifying the norms that really do enhance the prospects of profitable and stable cooperation—does power adaptive behaviour in its own right. First, consider the partner choice contexts we considered in Section 1: being good to seem good. In contexts of partner choice, the better you assess the moral facts, the better social interactions will go for you. You aim to choose, and be chosen by, partners who internalize not just any norms, but rather norms of cooperation, fair-dealing, trustworthiness, and commitment to their undertakings. You want such partners even if—perhaps especially if—they are locally unusual. To the extent partnership markets work in ways that defenders of partner choice models suppose, and to the extent that moral commitments are an important aspect of partner value in those markets, the commitments must be of a kind that motivate fair cooperation.

Second, while incentives to conform to any locally dominant norms are real, we should not think of agents as mere passive consumers of the local menu of norms. Agents influence their local normative environments. Most humans now live as globally invisible members of huge societies, but within these vast conglomerates, they live in sets of interconnected micro-worlds. They live in families, clubs, local workspaces, and informal social groups. Individual attitudes and actions can have significant positive and negative effects on these micro-worlds. Most of us will have experienced cooperative and friendly micro-worlds whose character was formed by the positive influence of a few key individuals. Less happily, most of us have also experienced micro-worlds whose cooperative dynamics have been ruined. Agents who accept, live, and promote prosocial, cooperation-sustaining norms (including the willingness to confront freeloaders) can influence these micro-worlds in ways that make them better for themselves (and others), and better for a wide range of particular plans. True moral beliefs are tools that can help an agent engineer their immediate social environment, even if their global

environment is impervious. No doubt the potential to change the local social world beneficially varies greatly from context to context. But we conjecture that it has often been present to some degree.

To sum up this article, then, in our view a version of reductive naturalism about moral norms can be built around one perspective on the evolutionary history of moral thinking. Moral truths are principles of action and interaction that support forms of cooperation and they are stable because they are fair enough to give almost everyone an incentive to continue to cooperate. In favourable cases, but only favourable cases, these norms are endorsed because they are true, and when endorsed, they support successful social interaction. The vindication is partial. For one thing, moral thinking is not just truth-tracking: it displays community membership and commitment to local mores, and norms solve coordination problems in ways that are independent of their truth. Moreover, to the extent that moral thinking is truth-tracking, it is error-prone. Our moral views are roughly analogous to the astronomical lore of the ancient world. Just as ancient astronomy was a response to the celestial world, moral views are a response to the opportunities and challenges of a world in which cooperation is profitable, but fraught with potentials for conflict, coordination failure, and misunderstanding. As in the case of pre-modern astronomy, these responses do not typically identify and solve those challenges ideally. But in a range of cases, the normative practices of individuals and groups are appropriately shaped by these challenges and the available solutions, and they enable individuals and groups to act adaptively in their social environments with some reliability. Moral thinking is neither a well-polished mirror of social nature, nor an adaptive fiction.

Acknowledgements

It is a pleasure to acknowledge the generous support from the Australian Research Council that made it possible for us to collaborate in writing this article. We also benefited from feedback from many people: two referees for this journal, Richard Joyce, Simon Keller, Dan Dennett and a cohort of his students; audiences at the ANU, Sydney University, Victoria University of Wellington, University of Cambridge, and St. Andrews.

Funding

Australian Research Council (FL130100141).

Kim Sterelny
Philosophy Program
Australian National University

Canberra, Australia,
kim.sterelny@anu.edu.au

Ben Fraser
Philosophy Program
Australian National University
Canberra, Australia
benjamin.fraser@anu.edu.au

References

Baumard, N., Andre, J.-B. and Sperber, D. [2013]: 'A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice', *Behavioral and Brain Science,* **36**, pp. 59–122.

Begossi, A., Hanazaki, N. and Ramos, R. [2014]: 'Food Chain and the Reasons for Fish Food Taboos among Amazonian and Atlantic Forest Fishers (Brazil)', *Ecological Applications*, **14**, pp. 1334–43.

Boehm, C. [2012]: *Moral Origins: The Evolution of Virtue, Altruism, and Shame*, New York: Basic Books.

Bowles, S. and Gintis, H. [2011]: *A Cooperative Species: Human Reciprocity and Its Evolution,* Princeton, NJ: Princeton University Press.

Boyd, R. [1988]: 'How to Be a Moral Realist', in G. Sayre-McCord (*ed.*), *Essays on Moral Realism,* New York: Cornell University Press, pp. 181–228.

Boyd, R. and Richerson, P. [1990]: 'Group Selection among Alternative Evolutionarily Stable Strategies', *Journal of Theoretical Biology,* **145**, pp. 331–42.

Boyd, R. and Richerson, P. [1992]: 'Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups', *Ethology and Sociobiology,* **13**, pp. 171–95.

Brink, D. [1989]: *Moral Realism and the Foundations of Ethics*, Cambridge: Cambirdge University Press.

Brosnan, K. [2011]: 'Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?', *Biology and Philosophy*, **26**, pp. 51–64.

Bulbulia, J. [2004a]: 'The Cognitive and Evolutionary Psychology of Religion', *Biology and Philosophy,* **19**, pp. 655–86.

Bulbulia, J. [2004b]: 'Religious Costs as Adaptations that Signal Altruistic Intention', *Evolution and Cognition,* **10**, pp. 19–38.

Chudek, M., Zhao, W. and Henrich, J. [2013]: 'Culture-Gene Coevolution, Large Scale Coooperation, and the Shaping of Human Social Psychology', in K. Sterelny, R. Joyce, B. Calcott and B. Fraser (*eds*), *Cooperation and Its Evolution*, Cambridge, MA: MIT Press, pp. 425–57.

Churchland, P. [1996]: 'The Neural Representation of the Social World', in L. May, M. Friedman and A. Clark (*eds*), *Minds and Morals*, Cambridge, MA: MIT Press, pp. 91–108.

Clarke-Doane, J. [2012]: 'Morality and Mathematics: The Evolutionary Challenge', *Ethics*, **122**, pp. 313–40.

Dennett, D. C. [1978]: *Brainstorms*, Montgomery: Bradford Books.

Dennett, D. C. [1991]: 'Real Patterns', *Journal of Philosophy*, **87**, pp. 27–51.

Dennett, D. C. [1995]: *Darwin's Dangerous Idea*, New York: Simon and Shuster.

Devitt, M. [1984]: *Realism and Truth*, Princeton, NJ: Princeton University Press.

Enoch, D. [2010]: 'The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It', *Philosophical Studies,* **148**, pp. 413–38.

Frank, R. [1988]: *Passion within Reason: The Strategic Role of the Emotions*, New York: W. W. Norton.

Fraser, B. [2012]: 'The Nature of Moral Judgments and the Extent of the Moral Domain', *Philosophical Explorations,* **15**, pp. 1–16.

Fraser, B. [2014]: 'Evolutionary Debunking Arguments and the Reliability of Moral Cognition', *Philosophical Studies*, **168**, 457–473.

Gauthier, D. [1987]: *Morals by Agreement*, New York: Oxford University Press.

Godfrey-Smith, P. [1996]: *Complexity and the Function of Mind in Nature*, Cambridge: Cambridge University Press.

Godfrey-Smith, P. [2012]: 'Signals, Icons, and Beliefs', in D. Ryder, J. Kingsbury and K. Williford (*eds*), *Millikan and Her Critics*, Oxford: Blackwell.

Harman, G. [1977]: *The Nature of Morality*, New York: Oxford University Press.

Harris, M. [1985]: *Good to Eat: Riddles of Food and Culture*, New York: Simon and Shuster.

Henrich, J. and Henrich, N. [2010]: 'The Evolution of Cultural Adaptations: Fijian Food Taboos Protect against Dangerous Marine Toxins', *Philosophical Proceedings of the Royal Society B*, **277**, pp. 3715–24.

Horwich, P. [1998]: *Truth*, New York: Oxford University Press.

Jackson, F. [1997]: *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Oxford University Press.

Joyce, R. [2006]: *Evolution of Morality*, Cambridge, MA: MIT Press.

Kitcher, P. [1993]: *The Advancement of Science*, New York: Oxford University Press.

Kitcher, P. [2011]: *The Ethical Project*, Cambridge, MA: Harvard Universty Press.

Lewis, D. [1969]: *Convention*, Oxford: Blackwell.

Lycan, W. G. [1990]: 'The Continuity of Levels of Nature', in W. G. Lycan and J. J. Prinz (*eds*), *Mind and Cognition*, Oxford: Blackwell, pp. 77–96.

Machery, E. and Mallon, R. [2010]: 'The Evolution of Morality', in J. Doris, F. Cushman and M. P. R. Group (*eds*), *The Moral Psychology Handbook*, Oxford: Oxford University Press, pp. 1–13.

Mackie, J. [1977]: *Ethics: Inventing Right and Wrong*, London: Penguin Books.

Mameli, M. [2001]: 'Mindreading, Mindshaping, and Evolution', *Biology and Philosophy,* **16**, pp. 595–626.

Nichols, S. [2004]: *Sentimental Rules: On the Natural Foundations of Moral Judgement,* New York: Oxford University Press.

Nichols, S. and Stich, S. [2004]: *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, Oxford: Oxford University Press.

Noë, R. [2001]: 'Biological Markets: Partner Choice as the Driving Force behind the Evolution of Cooperation', in R. Noë, J. van Hooff and P. Hammerstein (*eds*),

Economics in Nature: Social Dilemmas, Mate Choice, and Biological Markets, Cambridge: Cambridge University Press, pp. 93–118.

Ostrom, E. [1998]: 'A Behavioral Approach to the Rational Choice Theory of Collective Action', *American Political Science Review,* **92**, pp. 1–22.

Prinz, J. [2007]: *The Emotional Construction of Morals,* Oxford: Oxford University Press.

Putnam, H. [1975]: *Meaning and the Moral Sciences*, Cambridge, MA: Harvard University Press.

Richerson, P. and Boyd, R. [2001]: 'Institutional Evolution in the Holocene: The Rise of Complex Societies', *Proceedings of the British Academy,* **110**, pp. 197–234.

Richerson, P. and Henrich, J. [2012]: 'Tribal Social Instincts and the Cultural Evolution of Institutions to Solve Collective Action Problems', *Cliodynamics*, **3**, pp. 38–80.

Rocco, F. [2000]: *The Miraculous Fever-Tree: Malaria, Medicine, and the Cure that Changed the World*, London: Harper Collins.

Ross, D. [2006]: 'The Economic and Evolutionary Basis of Selves', *Cognitive Systems Research*, **7**, pp. 246–58.

Ruse, M. [1986]: *Taking Darwin Seriously*, Oxford: Blackwell.

Shafer-Landau, R. [2003]: *Moral Realism: A Defence*, Oxford: Oxford University Press.

Skyrms, B. [2010]: *Signals: Evolution, Learning, and Information*, Oxford: Oxford University Press.

Smart, J. J. C. [1959]: 'Sensations and Brain Processes', *Philosophical Review*, **68**, pp. 141–56.

Sterelny, K. [2003]: *Thought in a Hostile World,* New York: Blackwell.

Sterelny, K. [2007]: 'SNAFUS: An Evolutionary Perspective', *Biological Theory*, **2**, pp. 317–28.

Sterelny, K. [2010]: 'Moral Nativism: A Sceptical Response', *Mind and Language,* **25**, pp. 279–97.

Sterelny, K. [2012]: *The Evolved Apprentice,* Cambridge, MA: MIT Press.

Sterelny, K. [2014]: 'A Paleolithic Reciprocation Crisis: Symbols, Signals, and Norms', *Biological Theory*, **9**, 65–77.

Sterelny, K. [2016]: 'Cooperation, Culture, and Conflict', *British Journal for the Philosophy of Science*, **67**, 31–58.

Stich, S. [1993]: 'Moral Philosophy and Mental Representation', in M. Hechter, L. Nadel and R. Michod (*eds*), *The Origin of Values*, New York: Aldine de Gruyer, pp. 215–28.

Street, S. [2006]: 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies*, **127**, pp. 109–166.

Wilson, D. S. [2002]: *Darwin's Cathedral: Evolution, Religion, and the Nature of Society*, Chicago: University of Chicago Press.

Zawidzki, T. [2013]: *Mindshaping*, Cambridge, MA: MIT Press.