Moral Nativism: A Sceptical Response

Kim Sterelny
Philosophy Program
Victoria University of Wellington and The Australian National University

Draft of May 2007.

I. The Adapted Moraliser?

There are a number of universal and distinctive features of human life: features characteristic of all and only humans. Language is perhaps the most obvious of these. But our richly developed technologies and extensive networks of co-operation also distinguish us from other primates. We are distinctively talking, technological and co-operative apes. But we are also moralising apes. No other living primate moralises; arguably, all normal adult humans do. We think normatively, not just descriptively. The recent surge of interest in selectionist theories of human cognition and action has come to include a focus on moral cognition (see, for example, (Katz 2000; Joyce 2005; Hauser 2006) and on the adaptationist explanations of moral cognition. On the view being developed, moral cognition is a distinctive cognitive adaptation. Our ancestors varied genetically in their capacities for moral thought, and those who were capable of moral thought were fitter in virtue of that difference; presumably through greater access to the benefits of co-operation. Moral cognition, according to this suggestion, is as universal as language, and for the same reason: moral cognition depends on the evolution of a dedicated cognitive specialisation.

It is by no means easy to show that a trait is an adaptation. One strategy is to show that a trait is surprisingly apt for its environment; that there is a striking match between the phenotype of an individual organism and the adaptive demands on that organism. Optimality modelling is one way such matches can be identified. Such models can show the existence of a precise fit between (say) actual clutch size and the best trade-off between investment in current and in future reproduction. An optimality model of human moralising would estimate of the costs of moralising: the enemies you make; the costly acts you engage in to show you are not a hypocrite; the opportunities you must forgo. And it would include an estimate of the benefits: the potential sexual partners you impress as a reliable mate; the social partners you impress as a person of your word; the benefits you obtain by inducing others to act co-operatively. And on the basis of these, it would deliver an estimate of the optimal rate, circumstances and kind of moral action and injunction. Human behavioural ecologists could then test this prediction against observed human moral behaviour.

Nothing remotely like a quantitative behavioural ecological of moralising exists. In its place, we have a range of intuitive, informal approximations, together with an appeal to the distribution and development of moral cognition. These informal considerations basically connect the benefits of moralising to the benefits of enhanced co-operation. The benefits of co-operation are indeed great, and the idea is that we secure those benefits in part through normative cognition. We benefit from moral cognition both by co-operating ourselves and by inducing co-operation in others. Robert Frank and Richard Joyce have argued that in a world in which securing stable co-operation was increasingly important, our ancestral cognitive phenotypes would have left us with a weakness-of-will problem. Our primate minds would not have been able to resist the temptation to secure short-term benefit at the greater, but delayed, expense of fracturing trust (Frank 1988; Hauser 2006; Hauser, Young et al. forthcoming). Moral judgement increases the salience of the prosocial choice, making it easier for us to choose it. In support of this scenario Hauser 2006 notes that nonhuman animals do, indeed, find it difficult to defer reward (pp 343-346). Recent work on "strong reciprocation" shows that humans, in contrast to other primates, are psychologically prepared for co-operation. Public goods games show that many human enter social situations disposed to c-operate if they expect others to co-operate too (even if not co-operating serves their own interests better). And they are prepared to punish failures to co-operate at individual (so long as punishment is effective and not too expensive); see for example (Bowles and Gintis 2003; Fehr and Fischbacher 2003).

Human co-operative life is very distinctive; in this respect we are quite unlike other primates. So it is surely true that our psychology is selected for apt co-operation. That said, it is not possible for considerations of the kind Frank and Joyce develop to be compelling. The shift to an informal, intuitive argument makes it impossible to show that fit between phenotype and environment is precise and hence is of low probability. So developmental and distributional arguments are of especial importance in evaluating adaptationist hypotheses about human moral cognition. Marc Hauser (and his colleagues), John Mikhail and Susan Dwyer have recently developed these

arguments using an extended parallel between moral and linguistic cognition, arguing that we have a "moral grammar"[1].

II. The Grammar of Morality

Though not everyone is convinced[2], language is the most uncontroversial example of a special-purpose human cognitive adaptation. It is widely accepted that the human mind comes pre-equipped for the task of learning and using language. Language is independent of central, conscious cognitive processing. An agent does not have to decide to hear speech as language; it is automatic, mandatory. Agents parse sentences of their native language, recovering their organization, but they have no introspective access to the mechanisms which take speech as input and deliver to conscious awareness an interpretation of what has been said.. They recognise in aberrant cases the fact that there is something defective about sentence organizations. But agents do not have introspective access to the information about their native language that they use in such language processing. This information is tacit. It is not portable; it is not available to drive other cognitive processes, and hence cannot be expressed as assertions or beliefs. The conclusion: the normal development of the human mind results in a language module coming on stream: a special purpose subsystem that turns thought into talk and talk into thought. Hauser, Dwyer and Mikhail suggest that the human capacity for moral judgment has a structure parallel to that for language, both in development and operation.

The moral nativists argue that moral cognition is not just universal; it is universal in surprising and subtle ways. Thus Hauser 2006 argues that there is striking (though still only suggestive) evidence of cross-cultural uniformity in moral judgements. Most agents (though not all agents) draw an important moral distinction between acts and omissions; and they draw an important moral distinction between the foreseen but unintended consequences of actions, and those consequences that are both foreseen and intended. So in pursuit of the best overall outcome (for example: saving the most lives possible in the face of immanent diaster) it is morally acceptable to tolerate bad

---

[1] See (Dwyer 2006; Hauser 2006; Mikhail 2007; Hauser, Young et al. forthcoming; Hauser, Young et al. forthcoming) They credit the original analogy to John Rawls.
[2] For important sceptics, see (Cowie 1998; Tomasello 2003)

consequences that are foreseen but unintended, it is not morally acceptable to tolerate intended bad consequences; to intentionally do evil to block a still greater evil. So they judge in accordance with the so-called principle of double effect. In developing the grammatical model of morality, Hauser emphasises the subtlety and the apparent cultural invariance of these general principles which seem to determine particular judgements. But he also makes much of the introspective opacity of the principles which appear to guide specific judgements. While many agents judge in accordance with the principle of double affect, without help very few can consciously articulate that principle. So one crucial claim is that moral judgments are often fast, automatic, and systematically depend on tacit principles. Agents reliably make moral discriminations, regularly and predictably judging that some actions are morally appropriate and others are not. But they are typically unable to articulate the principles that guide these discriminations (Hauser, Young et al. forthcoming).

Hauser bases these claims on online survey work: subjects logged on to make judgments about moral dilemmas based around the famous trolley-bus thought experiments. Subjects read about scenarios in which an agent is faced between saving five lives at the cost of one. The cases differ in the kind of interventions which are necessary in order to save the five. In some interventions, the agent must directly kill the one (using him as a brake, buffer or obstruction). His death is the direct target of intervention. For example, to save the five from the runaway tram, an innocent bystander must be pushed onto the tracks to slow the tram, giving the five time to escape. In others, the one dies, but as a side-effect. The efficacy of the intervention to save the five does not in itself depend on its being fatal to the one. In one scenario, the agent has access to a switching mechanism that can divert the trolley bus from one line to another. Unfortunately, there is an innocent bystander on the spare line too, but while his death is foreseen it is not intended. The five would be saved if he were absent. Many of those making a call about these cases consistently discriminate between the two types of intervention. They judge that it is permissible to sacrifice the one as an unfortunate side-effect of saving the many, but it is not permissible to deliberately kill the one as the direct means of saving the many. However, few of those making this call can articulate the principles that divides the cases into two kinds. Their moral judgements are robust and replicable, but depend on principles that

are introspectively inaccessible. For this reason, moral development confronts a poverty-of-the-stimulus problem. We must be pre-wired for specific moral principles, for we come to rely on those principles in our fast and unreflective moral reasoning despite the fact that they have never been articulated and taught in the period in which moral competence develops. The moral nativists accept that there must be a significant cultural input to moral development, but they also argue that there is a very significant and specifically moral innate component.

In the rest of this paper I shall (first) suggest an alternative modular model of moral cognition; based on perception rather than language. For it offers a better model of the relationship between explicit and tacit moral beliefs. Even so (second) I shall argue against <u>any</u> modular model of moral cognition. Despite that scepticism, there are important aspects of moral cognition which really seem modular. So I shall offer (third) an alternative model of moral intuition and (fourth) an alternative account of our biological preadaptation for moral competence. That biological preparation, I shall suggest, is a hybrid of perceptual biases that make the effects of action salient to us, together with the informational engineering of our developmental environment. So this alternative owes a lot to recent sentimentalist views of moral cognition. The phenomena the moral grammarians have made central to their theory of moral cognition are real and important. But I think there is a better model of them than the one they offer.

I begin with the relationship between tacit and explicit moral cognition. Suppose Hauser and his co-workers are right to think that there is an abstract set of normative general principles that develop in every normal individual; principles whose general character is invariant but whose specific form depends on individual learning history; principles which automatically, rapidly, productively and unconsciously generate moral appraisals. Even if all this is true, there is still a striking difference between language and moral cognition. For in contrast to language, every normal agent also has conscious, and somewhat reflective moral principles. We endorse moral generalisations; we do not just make bullet-fast judgements about specific cases. What is the role of reflective morality, on the Hauser-Mikhail-Dywer picture?

One option would be to take reflective morality to be mostly unconnected from agents' actual practices or moral evaluation. It would be an epiphenomenon. This would be an awkward consequence of the linguistic model of moral cognition. Jon Haidt and his colleagues have argued that we are apt to overstate the importance of reflective morality and conscious moral reasoning: he and his colleagues argue that such conscious reasoning is often the post-hoc rationalisation of rapid, emotionally mediated responses. But they certainly do not argue that conscious moral reasoning is epiphenomenal; especially not when it is collective (see for example (Haidt and Bjorklund forthcoming). Haidt is surely right. While no doubt conscious moral reasoning is sometimes confabulation, it would be very implausible to claim that reflective morality is morally epiphenomenal. For agents seem to change moral practices (converting to vegetarianism, for instance) as a result of newly acquired reflective moralities. Moral vegetarianism and similar cases also make it implausible to suggest that that reflective morality is incomplete introspective access to the moral grammar (access of any kind, of course, would be an important contrast between grammar and moral grammar)[3]. For reflective morality is too unstable in individual agents, and too variable across agents. Agents convert to utilitarianism; rejecting the act/omission distinction and the doctrine of double affect. In contrast, once developed, a moral grammar is presumably fixed. So the moral grammarians seem committed to a dual process model of the mind (as in, for example, (Stanovich 2004)). We have both tacit and explicit moral cognition, and the systems interact. Yet none of the moral grammarians have a model of the interaction between the fast, tacit, automatic system and the slow conscious system. Syntax can offer no model of this interaction, because syntactic processing does not depend on conscious reasoning at all.

There is a second crucial difference between linguistic and moral cognition. It is not just general principles of language which are tacit, unavailable to introspection. Structural representations of particular utterances are also introspectively unavailable. In hearing a sentence, in some sense we must compute its organization or structure, for that organization plays a central role in sentence meaning. We understand sentences and to understand them, we must represent sentence structure. But in the

---

[3] For experimental work seems to show that agents have some conscious access to the principles that underlie their judgements: for example, they could articulate a principle appealing to the moral difference between acts and omissions but not the double effect principle. See (Cushman, Young et al. 2006)

standard case, we do not have syntactic beliefs. The parsing system takes as input sound from speech, and gives as output a representation of utterance structure. But that output does not consist of beliefs about sentence structure. In contrast, the output of a normative judgement system (if there is one) consists of beliefs about what should or should not be done. We may not always be able to articulate the general principles that guide our normative judgements. But we can certainly articulate those judgements themselves and make them public[4]. We do not just find ourselves inclined to act one way rather than another. Moreover, these judgements are thick rather than thin. Our culture provides us with a rich and nuanced set of tools for the moral evaluation of acts and agents; 'cruel", "capricious, "spiteful", "kind", "vengeful" and so on. . Law has developed a technical vocabulary and so has moral philosophy. But in contrast to the specialist vocabularies of linguistics, psychology or economic, both vocabularies are built on, and continuous with, the quite rich resources available to non-specialist agents.

That is a no accident. For one of the functions of moral judgement is the persuasion of others. We seek to persuade others to share our moral views both in particular cases and with respect to general principles. Haidt and Bjorklund quite rightly emphasis the fact that moral reasoning is often collective and social; it is a central part of social life (Haidt and Bjorklund forthcoming. Moralising is not a private vice. The contrast with language is important. There is no syntactic analogue of moral persuasion. In making normative appraisals, agents are not just in the business of guiding their own behaviour; they are in the business of guiding and evaluating the actions of others. This is a core feature of moral cognition. In contrast to syntactic judgement, moral judgement is in the public domain. We have no interest in the syntactic judgements of others: the more invisible syntactic processing is to the users of language, the better. It is like our retinal image. Artists excepted, we do not want to know about retinal images. Rather, we want to, and we normally do, "see through" the image to the world. But we have very good reasons to identity the normative judgements of others.

---

[4] Thus there is at least one important difference between acquiring grammar and acquiring the local morality. It is possible to argue that a specific logical problem confronts the language learning: they get no negative data. Ungrammatical sequences are not produced but labelled as deviant; they are just not produced at all. Hence we need to explain what blocks the acquisition of over-generalising grammars. That is not true of moral education: forbidden acts are both produced and described; and their forbidden character is made explicit. The principles of moral life may not be made explicit in moral education, but children are exposed to both positive and negative examples.

They are important indicators of how those others will act. And when we disagree, those views are the targets of our persuasive efforts. The role of moral judgement in human social life requires specific moral judgement (and perhaps the principles which guide those judgements) to be in the public domain. It is no accident that there is explicit moral thought.

III. Moral Vision.

One moral we might draw is that visual perception might be a better model of moral appraisal than grammar. For vision offers an account of the division of labour between tacit and conscious processing that we might export to moral cognition. As Fodor saw it, the visual modules deliver to central systems <u>a prima facie representation of how the world is</u>. These modular outputs are belief-like both in using representational categories that are also constituents of fully fledged beliefs, and in being introspectably accessible and explicit in belief-like ways. Vision tells the conscious mind how the world seems to be in terms the conscious mind itself uses, but without an agent having conscious access to the processes which transform the low-level detection of features of the sensory array into an estimate of the 3-D structure of the immediate environment. These deliverances of the visual module depend only on current visual input, and while we normally believe our eyes, we can overrule them, recognising illusions as illusions. Our all-things-considered judgement of how the world is not always congruent with how the world appears visually. There are visual illusions, but we are not always tricked by them, even when our eyes are (Fodor 1983)..

 Likewise, we might suppose that the moral appraisal subsystem delivers a judgement about the prima facie moral status of a particular action: a judgement that our central processor, with access to all we think, sometimes over-rules. The relation between vision and belief gives us a better model of the relationship between modular and conscious moral judgement. We might, for example, have relatively hard-wired automatic mechanisms that notice (by recognising others' emotions and feelings) harmful effects on those to whom we are affiliated. And we might have more or less automatic aversive responses to the sufferings such harms cause. This aversive

response could then be recruited as the basis of a prima facie evaluation system (as Prinz and Nichols have argued). The following is a good rule of thumb: acts that cause our intimates to suffer and which cause us to feel acutely uncomfortable in their commission are wrong. But as we know, sometimes one really does have to be cruel to be kind, and so it is only a good rule of thumb. We get moral illusions when the cues to moral evaluation to which we are architecturally committed are misleading. The moral module can be encapsulated without moral judgment itself being encapsulated. Moreover, moral modules might vary little across time and from culture to culture, without agents explicit, conscious, all-things-considered judgements being similarly constrained. Using the relation between vision and cognition as a model, we can see how it would be possible to automate prima facie moral evaluation. We could rely on a set of cues to the effects of actions and situations on the emotions of one's nearest and dearest. We could have a fast-and-dirty way of evaluating actions based on those emotional responses, for those responses are always relevant to the moral status of actions. Automated moral judgement could be tuned to internal cues of aversion.

While I think this is a more plausible view of a moral module, I still do not believe it. I do not think intuitive morality is largely invariant across culture. The recent history of the western world shows massive shifts in intuitive response, not just reflective morality. I grew up in an Australia in which homosexual acts caused moral disgust. In recent years there has been a shift in intuitive response to animal cruelty; the appropriate discipline of children; of the role of women in the world; on deference to social superiors, of being overweight. History shows considerable change in our visceral response to many situations: it is not so long since public execution was an immensely popular public entertainment in England. History and anthropology give us good reason to suppose that there is two-way interaction between reflective and intuitive morality, and not just in a critical period of childhood or early adolescence, when the child is learning the specific local form of the universal system of moral principles. Moreover, this is exactly what we should expect. For the moral problem is not a modular problem.

Vision is modular because vision is central to <u>online cognition</u>. It is central to avoiding obstacles and more generally, guiding ones motion through the world; it is central in monitoring the effects of ones own actions and in co-ordinating actions with others; in monitoring one's immediate physical and social environment. Most information that comes to us through vision has a short-shelf life. If it is relevant at all, it is relevant now. So the system has to be fast. It can be encapsulated because the visual discriminations we make typically depend on very general and very stable features of our physical environment. Smooth change in apparent size, for example, is almost always the result of motion, not actual change in the object in view. Interpreting speech is also a job for online cognition. Since we cannot store mere speech sound patterns in long-term memory, if we are to understand others, we must understand them as they talk. So the adaptive rationale of perceptual modules depends on two features: first: the tasks which they support are urgent, and second: the features of environments on which their reliability depends are stable.

Moral decision-making, in contrast, typically does not have this urgency: we very rarely have to judge and act within a fraction of a second. Our moral evaluations of actions and situations are typically aspects of planning, of offline cognition than online cognition. We evaluate scenarios in conversation and imagination; we do not just evaluate actions and circumstances we perceive, and when we do, we rarely have to act in milliseconds. We imaginatively project ourselves into possible futures and pasts, in "mental time travel" (Suddendorf and Busby 2003). We can use the resources of episodic memory to construct sensory representations of situations, even in the absence of perceptual input. We can decontextualise our representations of situations, stripping away detail that is irrelevant. We can use the resources of declarative memory to represent situations and reason about them (Gerrans forthcoming). These modes of offline cognition are all important to moral thinking. Even the reactive, intuitive systems of moral evaluation are engaged in advance, in imagining future plans as well as retrospectively judging past ones. Hauser's own trolley-bus experiments are not representative of everyday decision making. But they do illustrate the role of offline cognition in moral appraisal. In that respect, these are typical instances of moral cognition. If, though, moral intuitions are not the output of a system of innate tacit principles, where do they come from?

IV. Intuitive Moral Judgement

Moral assessment is fast and automatic. We do not have to decide whether to evaluate a situation normatively when we read of the abduction of a three-year old girl. Moreover, there does seem reason to suspect that normative thought is a universal feature of human society and of normal agency; and also reason to suspect that there are at least family resemblance relations between the moral systems of different cultures. The capacity for moral cognition seems to develop impressively early and with impressive robustness. Consider, in contrast, literacy and numeracy. Their acquisition requires require extensive formal education, yet despite that formal education the development of these skills is still fragile. Those who are sceptical of the moral grammarians' picture will need to explain why moral cognition seems modular, even though it is not.

Natural minds are good at learning to recognise patterns — similarities between instances — and the exercise of that capacity results in intuitive judgements about new cases. Pattern recognition is productive: it extends to indefinitely many new cases. Such judgements are fast, automatic, and often the agent cannot explain the basis of the judgement. In a series of vivid examples, Chomsky argued that grammar could <u>not</u> be captured as set of simple lexical patterns. To capture (say) the relationship between indicative and yes-no questions, we need to identify the abstract structure — the organization of lexical items into subsentential constituents. Chomsky's argument is very important, because if sound, it shows that syntax is not pattern recognition. Agents do not learn their language by generalising from familiar cases; using familiar examples as models for new constructions (though see Tomasello 2003 for a view of language acquisition appealing to such mechanisms). No Chomsky-style argument has been run for moral cognition. For while Hauser has certainly argued that even elementary moral judgements depend on subtle facts about situations, the facts in question are not moral but intentional. They are facts about goals, intentions, consequences (Hauser 2006). So a hypothesis: our intuitive moral judgements are generalisations from exemplars. Kind and generous acts are those that

resemble paradigmatic moments of kindness or generosity, and so on for other evaluations.

Pattern recognition is fast and automatic, once the abilities have come on line. Chess experts, for example, assess positions rapidly and accurately, and it is very hard for a chess player not to see a chess position as a position. An expert cannot but see a knight on E6 as a knight dominating the centre, even if he can also see it as a shape on a pattern of squares. Moreover the metrics underlying pattern recognition are often tacit. Expert birders, for example, can often recognise a bird from a fleeting glimpse without being able to say how they recognise it. It has the "jizz" they will say, of a brown falcon. So if intuitive moral judgements are the result of pattern recognition capacities, it is no surprise that they have the rapidity and introspective translucence that Hauser and his colleagues have identified. Furthermore, an exemplar-based view of moral judgement is independently plausible. For one thing, at least in the western world, moral education is largely-example based. Children are exposed to general principles (though if the moral grammarians are right, not to some of the ones they actually use). But they are also exposed to a rich stock of exemplars. The narrative life of a community — the stock of stories, songs, myths, and tales to which children are exposed — is full of information about what actions are to be admired, and which are to be deplored. Young children's stories include many moral fables: stories of virtue, of right action and motivation rewarded; of vice punished. So their narrative world is richly populated with moral examples.

So too, for many, is their world of individual experience. The moral grammarians (in particular, Susan Dwyer) sometimes write as if the task of children was to induce moral norms by mere observation of regularities in the behaviour of local adults. But children do not just look and listen. They act. In their interactions with peers, they encounter many morally charged situations, especially those to do with harms and with fairness. So there are lots of particular experiences annotated with their moral status to act as input to a pattern-recognition learning system (a model defended in (Churchland 1996).

Moreover, an exemplar-based view of moral intuition makes sense of an otherwise surprising fact: we judge harms by inaction less severely than harms of direct action. From an adaptationist perspective, the act/omission distinction is puzzling and arbitrary. My defection by inaction is just as costly to my partner or my group as a defecting act would be. Failing to help a partner in dire need dooms them as certainly as a malicious act would. If the adapted function of morality is to support and extend prosocial behaviour, we would expect omissions and commissions to be morally equivalent. Moreover, omissions are more difficult to detect. Passive deception — allowing you to act on a misapprehension that I recognise and could correct — will typically be more difficult to identify than a deliberate lie. The same is true of failures to provide material aid: it will often not be obvious to others whether the omission was deliberate, whether the defecting agent was aware of the situation and in a position to intervene. This intensifies the puzzle. Whether it is in my interest to defect depends on the reward of successful defection, the risk that the defection will be detected together with the severity of punishment, if detected. To keep a constant level of deterrence, as the risk of detection goes down, the severity of the punishment should increase (which is presumably why so many crimes were capital in our legal system before the organization of a police force). The well-adapted intuitive moraliser should be incandescent with rage at cryptic harming by omission, but somewhat more forgiving of upfront, in your face, fuck-you defection.

The puzzle abates if intuitive judgement depends on generalisation from a set of exemplars. For the detection problem above will bias the training set in favour of acts of commission. Children's theory of mind skills are not yet fully on line, so the obvious, unmistakable examples of kindness or cruelty; of fairness or greed, will be things we do, not things we fail to do. So the cases around which our similarity measures coalesce will be a core of positive interventions. We judge omissions less harshly because they are further from our paradigms of the punishable. Sunstein argues that moral cognition often depends on moral heuristics, and he takes the act/omission distinction to be such a heuristic (Sunstein 2005). I agree, but the model sketched here explains the attraction of the heuristic: it is based on an epistemic bias in favour of positive cases. If this is right, if there are species of social interactions in which omissions are as obvious and as salient as commissions, then in those

interactions, the act/omission heuristic should disappear. My armchair social psychology is no evidence, but I suspect bargains and promises are such a species. Failing to keep a promise or return a favour is a salient omission, and one not mitigated by being a failure to act.

V. The Biological Preparation of Moral Development

There are two strands to the moral grammarians conception of moral cognition. First, they take moral judgement to depend on general, abstract moral principles, though principles that will typically be initialised to particular contexts and cultures (using local specifications of harm, equity, kin, and so forth). Second, they think acquiring moral cognition is biologically prepared. We are adapted for moral thought in such a way that its development is accelerated and made more reliable. While we genuinely learn about our moral universe, we can learn about that universe more rapidly and reliably because our minds are specifically adapted for this learning task. In turn, those adaptations for moral learning damp-down cross cultural variation in moral systems. While sceptical about the appeal to tacit principles, I agree that moral development is robust because we are biologically prepared for moral education. But I think that preparation consists in the organization of our developmental environment and through specific perceptual sensitivity.

Nativists typically describe learning as if it were the passive observation of information that happens to be available in the environment. That overlooks two forms of epistemic niche construction. First, parents do not just wait and hope that children will acquire the information they need. They organise the informational world of their children. They make trial and error learning safer, by substituting social signals of failure for environmental ones. They provide informational resources: toys, tools and other props. They intervene in trial and error experiments by children: much human learning is a powerful hybrid of individual and social learning (Sterelny forthcoming d). And, of course, there is a significant amount of explicit teaching. As a consequence of these interventions, the parental generation actively engineers the learning environment of the next generation: trial-and-error learning takes place in an enriched and highly structured environment. The reliability of the social transmission

of information across the generations is much enhanced by parents' informational engineering. Second, children themselves are far from passive: they are themselves active epistemic agents. They do not wait for the environment to provide them with the information they need: they actively probe their physical and social worlds (Sterelny 2003). These general points about human social learning apply to the acquisition of moral cognition in children. Children do not acquire information about the moral opinions of their community just by observation of adult practice: of what adults do and avoid doing. Adults talk as well as act: both with one another and with children. They respond with emotional signals to what others say and do, and those responses carry normatively important information. Moreover, children experiment with and manipulate their surrounds. Children's social worlds are full of disputed terrain, especially to do with issues of fair division and responsibility. Children collide with the moral views of their peers, and they attempt to impose their own views on those peers. Few children could learn from the norms of fair division from the untrammelled actions of their brothers and sisters. But they have a good chance of learning them from overhearing and taking part in discussions over how spoils are to be divided; "you cut; I choose" and similar division rules.

So a child's moral development normally takes place in a prepared environment. But we are also prepared for moral education by the phenomena we find perceptually salient. Nativists often model the situation of the first language learner as that of the radical interpreter: a hypothetical anthropologist in an alien community, facing the task of coming to interpret the alien language. In the case of language, this model may well be appropriate for there is no interesting distinction between being able to fully describe a language and being able to speak it. So we can think of the child's language learning task as structurally similar to that of the radical linguist. In the moral domain, the ethnographer's task is also descriptive: namely to describe the norms of a village or community. But there is a difference between being able to describe moral norms and sharing those norms; of making them one's own norms. The task of the child is not to discover the set of practices governing a particular community. Rather, her task is to join her community; to share rather than describe those norms. She can do so because moral cognition is embodied. Her own emotional responses, and her sensitivity to the responses of those that surround her, play a

crucial scaffolding role in her entering the community in which she lives. In learning a new language — the radical linguistics task — it helps immensely to already have a language. Language is a wonderful cognitive tool. In coming to share a set of norms — initially, in developing a reactive morality — in would not help to begin with a different set of norms, an inconsistent set of reactions and evaluations. The anthropologist's project is not a helpful model of the child's project.

Thus I align myself with the Humean revival: we are perceptually tuned to the emotions and emotional responses of others, and to our own emotional responses to others. We typically notice other's distress, and we do not just notice it, we respond with emotions of our own. Those emotions, too, we notice by an internal analogue of perception. Such emotional contagion and its relatives is take-off point for the Humean revival in psychology and philosophy (Prinz, Nichols, Haidt, Pizarro[5]) known as sentimentalism. We respond positively to kindness; we are aware of our own positive response, and we convert that visceral reaction into a normative judgement. Moral norms are grafted on top of our dispositions to respond emotionally in characteristic ways to stereotypic stimuli: for example, with disgust to body products and rotting food; with distress/discomfort to the suffering of others (or at least those in our group). These emotional responses mark certain types of events; they are salient. As Nichols sees it, putative norms which line-up with these characteristic responses are more stable. They are more likely to be endorsed, respected and taught that arbitrary norms (Nichols forthcoming).

Moreover, such norms are much more learnable. Entrenched, automatic responses make certain situations and actions salient to children: normal children, for example, notice when their play-mates are distressed. Their own emotional responses to the emotions they notice are motivating. Distress makes most children uncomfortable. Our emotions make certain types of action and situation salient: we notice the emotions and reactions of others, and that in itself will narrow moral search space: in

---

[5] See for example: (Pizarro 2000; Nichols 2004; Pizarro, Detweiler-Bedell et al. 2006; Haidt and Bjorklund forthcoming; Nichols forthcoming; Prinz forthcoming; Prinz forthcoming). My own views place a greater weight on the role of explicit moral thought, and give an explicit and central role to informationally organised developmental environments. But these are just differences in emphasis. See (Clark 2000) for a view similar to mine on the interaction between tacit and explicit knowledge; similar too on the basis of the implicit system.

thinking about the content of moral norms, they will tend to think of human emotions, reactions, and the stimuli that engage them. Children (and to adults) notice these phenomena, and are motived by them, independently of, and prior to moral education. So a child's evidential base includes not just patterns in others' actions; it includes information about others' emotional responses to those actions, and information about that agent's own emotional response (Scott forthcoming).

On this picture, then, we are biologically prepared to develop moral cognition, but not because our minds are adapted to moral concepts and principles, but because moral cognition is a natural development of our existing emotional, intellectual and social repertoire. The idea is a familiar response to linguistic nativism. Perhaps our minds are not adapted to use language by having language-specific information built into us. Rather, languages are adapted to be used by agents with minds like ours, and their structural similarities follow from that; see for example (Deacon 1997). While intriguing, no-one has succeeded in developing this idea into a systematic theory of the organization of language. But the parallel suggestion about moral cognition is very compelling. Our suite of emotional reactions — especially those concerned with reciprocation, sympathy, empathy disgust and esteem — shape and constrain (though certainly not determine) moral cognition. On this view, moral cognition develops from an interaction between emotions, exemplar-guided intuitions and explicit principles. We might expect similar developmental trajectories early, because at that stage of the development of moral cognition, emotional response helps select the exemplars.

It is important to recognise a role for explicit principles. Mapping the relationships between gut-feeling morality, explicit moral reasoning, and emotion is a challenge to every model of moral cognition. While I am sympathetic to the sentimentalist family of views, its defenders sometime underplay the role of explicit moral reasoning. For example, (Haidt and Bjorklund forthcoming) defend a version of the Humean picture in which intuitive moral judgement is the reflection of our emotional responses of aversion and approval, and in which conscious moral reasoning is typically a post-hoc response to intuitive judgement. They do not think that conscious moral reasoning is always confabulation. They accept that (especially in social interaction) conscious

reasoning can determine moral judgement. But typically judgement is a result of emotional response.

Hauser rightly think that this is too bottom-up a picture of judgement. He discusses Rozin's work showing that the disgust moral vegetarians typically come to feel for meat is a consequence rather than a cause of their moral convictions. There is, for example, no reason to believe that moral vegetarian views are primed by a prior hypersensitivity to disgust reactions. So while emotion is certainly intimately linked to moral appraisal, Hauser points out that it does not follow that the emotions drive the appraisals. Indeed, he suggests that the explanatory arrow runs the other way. As he points out, situations must be appraised; actions recognised for what they morally are, before emotions can be engaged. Emotional response is not typically dumb and reflexive. While that is right, it is not obvious that the analysis of a situation that precedes moral response must include a <u>moral</u> as well as a <u>intentional</u> evaluation. Moreover, while the moral vegetarianism example illustrates a diachronic interaction between moral emotions and moral principles, showing that over time, principles can shape emotional responses, this example does not illustrate the operation of *tacit moral principles*. The principles that fuel moral vegetarianism are typically conscious, explicit, articulated. I still do not see an argument here for tacit and abstract general principles. But his broader point is surely right: we do need to identify the ways top-down reasoning effects emotions, exemplars, and, specially, generalisation from exemplars. A key form of moral argument is to try to induce another to see similarity relations between cases they already evaluate morally in a distinctive way and the case under dispute. And these arguments sometimes work.

VI Concluding Summary

In brief, the picture presented here does accept that the development of moral cognition is supported by our evolved biology, but it does not invoke domain-specific principles on the model of grammar. That might be too minimal a view of our biological endowment. I have argued that children develop their moral intuitions about, for example, fairness on the basis of prototypical representations of fair and unfair actions. Prima facie moral judgement might well be based on pattern

recognition guided by paradigm instances. So given that a child has the concept of a moral transgression, perhaps it is not hard to work out which acts count as transgressions. But perhaps acquiring that core concept is the crucial challenge to a non-nativist view of morality[6]. Richard Joyce argues that the concept of a moral norm itself poses they key learning problem. He writes: "it is puzzling what there even could be in the environment — even a rich and varied environment — for a generalized learning mechanism to latch onto to develop the idea of a *moral transgression*. … Administer as many punitive beatings as you like to a general-purpose intelligence; at what point does this alone teach it the idea of a moral transgression?" (Joyce 2005) p137)

In the absence of an explicit account of moral concepts and of what is required to learn them, it would be premature to reject this minimal but still domain specific nativism of Joyce. That said, I remain to be convinced. For one thing, there is a tension between accepting an adaptationist model of evolved moral cognition (as Joyce does) and arguing that moral concepts are strictly unlearnable by a trial and error process. For both are incremental, feedback-driven hill-climbing models of the construction of a capacity. One mechanisms operates within a generation, and the other over many generations. There is a structural similarity between learning and adaptive evolution: they are both trajectories of gradual improvement in response to success and failure signals from the environment. If the concept of a moral transgression is <u>strictly unlearnable</u> from experience by a mind without any moral concept, it is cannot evolve via a Darwinian trajectory from such a mind, either.

---

[6] Dwyer, if I understand her, thinks this. She is impressed by the fact that children have command of the concept of a moral norm, distinguishing it from a merely conventional norm, very early in life, and so she thinks children must have this concept built-in. It is not clear that this datum is holding up to further experiment (Kelly, Stich et al. 2007), but even if it does, I am less impressed by it than is Dwyer, for I think children get plenty of cues about the differences between conventional and moral norms: young children catch on to harm norms, and violations of these result in distinctive emotional signals from the victims (and hence in the onlookers). Moreover, even if parents care about conventional norm violations as intensely as they care about moral norm violations, they offer different justifications and explanations of the two ((Scott forthcoming). Moreover, games can also play a role in the acquisition of normative cognition, teaching about turn-taking and the like, as well as the distinction between an authority-dependent rule (as in the rules of a game, which can be changed by negotiation) and authority-independent rules.

Even so, a minimal nativism of the kind Joyce envisages is clear possible and compatible with most of what I have said. My main point is that nativism, and especially richer views which attribute significant domain specific information, are not mandatory. Normative cognition could be genuinely universal, and develop robustly without that development relying on innate, tacit, abstract principles. Many human competences develop reliably in the downstream generation because the upstream generation structures downstream learning environments to support those developments. The development of a cognitive competence is often stabilised by ecological engineering, rather than a genetically-based special purpose cognitive mechanism. This general view may well be true of moral competence: it is transmitted culturally yet reliably. The parental generation engineers the informational environment in which the next generation develops, thus guaranteeing the development of moral competence.

Barkow, J. H., L. Cosmides, et al., Eds. (1992). The Adapted Mind: Evolutionary Psychology and the Generation of Culture. Oxford, Oxford University Press.

Bowles, S. and H. Gintis (2003). Origins of Human Cooperation. Genetic and Cultural Evolution of Cooperation. P. Hammerstein. Cambridge, MIT Press: 429-443.

Cowie, F. (1998). What's Within? Nativism Reconsidered. Oxford, Oxford University Press.

Cushman, F., L. Young, et al. (2006). "The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm." Psychological Science 17(2): 1082-1089.

Dwyer, S. (2006). How Good Is The Linguistic Analogy? The Innate Mind: Volume 2 Culture and Cognition. P. Carruthers, S. Laurence and S. Stich. Oxford, Oxford University Press: 237-255.

Dwyer, S. (forthcoming). How Good Is The Linguistic Analogy? The Innate Mind: Culture and Cognition. S. Stich, S. Laurence and P. Carruthers. Cambridge, Cambridge University Press.

Fehr, E. and U. Fischbacher (2003). "The Nature of Human Altruism." Nature 425: 785-791.

Fodor, J. A. (1983). The Modularity of Mind. Cambridge, MIT Press.

Frank, R. (1988). Passion Within Reason: The Strategic Role of the Emotions. New York, WW Norton.

Gerrans, P. (forthcoming). "Mechanisms of Madness: Evolutionary Psychiatry Without Evolutionary Psychology." Biology and Philosophy.

Haidt, J. and F. Bjorklund (forthcoming). Social Intuitionists Answer Six Questions about Moral Psychology. Moral Psychology. W. Sinnott-Armstrong. Oxford, Oxford University Press. 3.

Hauser, M. (2006). Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong. New York, HarperCollins.

Hauser, M., L. Young, et al. (forthcoming). On Misreading the Linguistic Analogy: Response to Jesse Prinz and Ron Mallon. Moral Psychology and Biology. W. Sinnott-Armstrong. Oxford, Oxford University Press.

Hauser, M., L. Young, et al. (forthcoming). Reviving Rawl's Linguistic Analogy. Moral Psychology and Biology. W. Sinnott-Armstrong. Oxford, Oxford University Press.

Joyce, R. (2005). Evolution of Morality. Cambridge, Mass, MIT Press.

Joyce, R. (forthcoming). Is Human Morality Innate. The Innate Mind: Culture and Cognition. S. Stich, P. Carruthers and S. Laurence. Cambridge, Cambridge University Press.

Katz, L., Ed. (2000). Evolutionary Origins of Morality: Cross-Disciplinary Perspectives. Exeter, Imprint Academic.

Mikhail, J. (2007). "Universal Moral Grammar: Theory, Evidence, and the Future." Trends in Cognitive Science.

Nichols, S. (forthcoming). Innateness and Moral Psychology. The Innate Mind: Structure and Content. P. Carruthers and S. Laurence. New York, Oxford University Press.

Stanovich, K. (2004). The Robot's Rebellion. Chicago, University of Chicago Press.

Sterelny, K. (2003). Thought in a Hostile World. New York, Blackwell.

Sterelny, K. (forthcoming d). "The Evolution and Evolvability of Culture." Mind and Language.

Suddendorf, T. and J. Busby (2003). "Mental Time Travel In Animals." Trends in Cognitive Science 7(9): 391-397.

Tomasello, M. (2003). Constructing A Language: A Usage Based Theory of Language Acquisition. Cambridge, Harvard University Press.

Bowles, S. and H. Gintis (2003). Origins of Human Cooperation. Genetic and Cultural Evolution of Cooperation. P. Hammerstein. Cambridge, MIT Press: 429-443.

Churchland, P. (1996). The Neural Representation of the Social World. Minds and Morals. L. May, M. Friedman and A. Clark. Cambridge, MIT Press: 91-108.

Clark, A. (2000). Word and Action: Reconciling Rules and Know-How in Moral Cognition. Moral Epistemology Naturalised R. Campbell and B. Hunter. Alberta, Canada, University of Calgary Press: 267-290.

Cowie, F. (1998). What's Within? Nativism Reconsidered. Oxford, Oxford University Press.

Cushman, F., L. Young, et al. (2006). "The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm." Psychological Science 17(2): 1082-1089.

Deacon, T. (1997). The Symbolic Species: The Co-evolution of Language and the Brain. New York, W.W Norton.

Dwyer, S. (2006). How Good Is The Linguistic Analogy? The Innate Mind: Volume 2 Culture and Cognition. P. Carruthers, S. Laurence and S. Stich. Oxford, Oxford University Press: 237-255.

Fehr, E. and U. Fischbacher (2003). "The Nature of Human Altruism." Nature 425: 785-791.

Fodor, J. A. (1983). The Modularity of Mind. Cambridge, MIT Press.

Frank, R. (1988). Passion Within Reason: The Strategic Role of the Emotions. New York, WW Norton.

Gerrans, P. (forthcoming). "Mechanisms of Madness: Evolutionary Psychiatry Without Evolutionary Psychology." Biology and Philosophy.

Haidt, J. and F. Bjorklund (forthcoming). Social Intuitionists Answer Six Questions about Moral Psychology. Moral Psychology. W. Sinnott-Armstrong. Oxford, Oxford University Press. **3**.

Hauser, M. (2006). Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong. New York, HarperCollins.

Hauser, M., L. Young, et al. (forthcoming). On Misreading the Linguistic Analogy: Response to Jesse Prinz and Ron Mallon. Moral Psychology and Biology. W. Sinnott-Armstrong. Oxford, Oxford University Press.

Hauser, M., L. Young, et al. (forthcoming). Reviving Rawl's Linguistic Analogy. Moral Psychology and Biology. W. Sinnott-Armstrong. Oxford, Oxford University Press.

Joyce, R. (2005). Evolution of Morality. Cambridge, Mass, MIT Press.

Katz, L., Ed. (2000). Evolutionary Origins of Morality: Cross-Disciplinary Perspectives. Exeter, Imprint Academic.

Kelly, D., S. Stich, et al. (2007). "Harm, Affect, and the Moral/Conventional Distinction." Mind and Language **22**: 117-131.

Mikhail, J. (2007). "Universal Moral Grammar: Theory, Evidence, and the Future." Trends in Cognitive Science.

Nichols, S. (2004). Sentimental Rules: On the Natural Foundations of Moral Judgment
. New York, Oxford University Press.

Nichols, S. (forthcoming). Innateness and Moral Psychology. The Innate Mind: Structure and Content. P. Carruthers and S. Laurence. New York, Oxford University Press.

Pizarro, D. (2000). "Nothing More Than Feelings? The Role of Emotions in Moral Judgement." Journal for the Theory of Social Behaviour **30**(4): 355-375.

Pizarro, D., B. Detweiler-Bedell, et al. (2006). The Creativity of Everyday Moral Reasoning: Empathy, Disgust and Moral Persuasion Creativity and Reason in Cognitive Development. J. Kaufman and J. Baer. Cambridge, Cambridge University Press.

Prinz, J. (forthcoming). Against Moral Nativism. Stich and His Critics. D. Murphy and M. Bishop. Oxford, Backwell.

Prinz, J. (forthcoming). The Emotional Construction of Morals. Oxford, Oxford University Press.

Scott, T. (forthcoming). The Richness of the Moral Stimulus. Philosophy. Wellington, Victoria University of Wellington.

Stanovich, K. (2004). The Robot's Rebellion. Chicago, University of Chicago Press.

Sterelny, K. (2003). Thought in a Hostile World. New York, Blackwell.

Sterelny, K. (forthcoming d). "The Evolution and Evolvability of Culture." Mind and Language.

Suddendorf, T. and J. Busby (2003). "Mental Time Travel In Animals." Trends in Cognitive Science **7**(9): 391-397.

Sunstein, C. (2005). "Moral Heuristics." Behavioral and Brain Sciences **28**: 531-573.

Tomasello, M. (2003). Constructing A Language: A Usage Based Theory of Language Acquisition. Cambridge, Harvard University Press.