

The liver and the moral organ

Marc D. Hauser

Departments of Psychology, Organismic & Evolutionary Biology, and Human Evolutionary Biology Harvard University, Cambridge, MA 02138, USA

Drawing on an analogy to language, I argue that a suite of novel questions emerge when we consider our moral faculty in a similar light. In particular, I suggest the possibility that our moral judgments are derived from unconscious, intuitive processes that operate over the causal-intentional structure of actions and their consequences. On this model, we are endowed with a moral faculty that generates judgments about permissible and forbidden actions prior to the involvement of our emotions and systems of conscious, rational deliberation. This framing of the problem sets up specific predictions about the role of particular neural structures and psychological processes in the generation of moral judgments as well as in the generation of moral behavior. I sketch the details of these predictions and point to relevant data that speak to the validity of thinking of our moral intuitions as grounded in a moral organ.

Keywords: moral organ; linguistic analogy; moral judgments; cognitive faculties

A close friend of mine just went through a harrowing experience: his brother, who was suffering from a rare form of liver cancer, was in the queue for a liver transplant. Livers, like many other organs, are in high demand these days, and those in the queue are desperate for what will most likely turn into a life-saving operation. The fortunate thing about liver transplants is that, assuming the damaged cells have been contained to the liver, the operation is like swapping out a hard drive from your computer: plug and play. Sadly, neither our brains in toto, nor their component parts, are similarly constituted and even if they were, there would be a fundamental asymmetry in the swapping. I would be happy to receive anyone's liver assuming the donor's was healthy and mine not. I would not be happy to receive anyone's brain or brain parts, even if healthy. Though we can readily define regions of brain space, specify general functionality and describe wiring diagrams to other bits of neural territory—precisely the kind of descriptive information we provide for the liver, heart, eye and ear—the notion of 'organ' for the brain is more metaphorical than anatomical. But metaphors can be useful if we are careful.

Here, I would like to push the idea that we are endowed with a moral organ, akin to the language organ. The link to language is essential to the arguments I will develop here (Rawls, 1971; Harman, 1999; Dwyer 1999, 2004; Mikhail, 2000, in press), and have developed more completely elsewhere (Hauser, 2006; Hauser *et al.*, in press b). I therefore start with the argument that as a promising research strategy, we should think about our moral psychology in the way that linguists in the generative

tradition have thought about language. I then use this argument to sketch the empirical landscape, and in particular, the kind of empirical playground that emerges for cognitive neuroscientists interested in the neural circuits involved in generating moral judgments. I follow with a series of recent findings that bear on the proposed thesis that we have evolved a moral organ, focusing in particular on studies of patient populations with selective brain deficits. Finally, I return to the metaphor of the moral organ and point to problems and future directions.

FACULTIES OF LANGUAGE AND MORALITY

In a nutshell, when the generative grammar tradition took off in the 1950s with Chomsky's (1957, 1986) proposals, linguistics was transported from its disciplinary home in the humanities to a new home in the natural sciences. I am, of course, exaggerating here because many within and outside linguistics resisted this move, and continue to do so today. But there were many converts and one of the reasons for conversion was that the new proposals promised to bring exciting insights into the neurobiological, psychological, developmental and evolutionary aspects of language. And 50 years later, we are witnessing many of the fruits of this approach. This is true even though the theories and approaches to the biology of language have grown, with controversies brewing at all levels, including questions concerning the autonomy of syntax, the details of the child's starting state, the parallels with other organisms, and relevant to the current discussion, the specificity of the faculty itself. For my own admittedly biased interests, the revolution in modern linguistics carried forward a series of questions and problems that any scholar interested in the nature of things mental must take seriously. Put starkly, Chomsky and those following in the tradition he sketched, posed a set of questions

Received 1 August 2006; Accepted 1 September 2006

Correspondence should be addressed to Marc D. Hauser, Departments of Psychology, Organismic & Evolutionary Biology, and Human Evolutionary Biology Harvard University, Cambridge, MA 02138, USA.
E-mail: mdh@wjh.harvard.edu.

concerning the nature of knowledge and its acquisition that are as relevant to language as they are to mathematics, music and morality. I take the critical set of questions to include, minimally, the following five, spelled out in terms of the general problem of ‘For any given domain of knowledge . . .’

- (i) what are the operative principles that capture the mature state of competence?
- (ii) how are the operative principles acquired?
- (iii) how are the principles deployed in performance?
- (iv) are the principles derived from domain-specific or general capacities?
- (v) how did the operative principles evolve?

Much could be said about each of these, but the critical bits here are as follows. We want to distinguish between competence and performance, ask about the child’s starting state and the extent to which the system matures or grows independently of variation in the relevant experiential input, specify the systems involved in generating some kind of behavioral response, determine whether the mechanisms subserving a given domain of knowledge are particular to that domain or more generally shared, and by means of the comparative method, establish the evolutionary phylogeny of the trait as well as its adaptive significance. This is no small task, and it has yet to be achieved for any domain of knowledge, including language. Characterizing the gaps in our knowledge provides an essential road map for the future.

The linguistic analogy, as initially discussed by Rawls (1951, 1963, 1971), and subsequently revived by the philosophers Harman (1999), Dwyer (1999, 2004), and Mikhail (2000, in press), can be formalized as follows. We are endowed with a moral faculty that operates over the causal-intentional properties of actions and events as they connect to particular consequences (Hauser, 2006). We posit a theory of universal moral grammar which consists of the principles and parameters that are part and parcel of this biological endowment. Our universal moral grammar provides a toolkit for building possible moral systems. Which particular moral system emerges reflects details of the local environment or culture, and a process of environmental pruning whereby particular parameters are selected and set early in development. Once the parameters are set for a particular moral system, acquiring a second one later in life—becoming functionally *bimoral*—is as difficult and different as the acquisition of Chinese is for a native English speaker.

Surprisingly perhaps, though the theoretical plausibility of an analogy to language has been in the air for some time now, empirical evidence to support or refute this possibility has been slow in coming. In the last few years, however, the issues have been more formally stated, allowing the modest conclusion that there are new questions and results on the table that support the heuristically useful nature of this analogy; it is too early to say whether there is a deeper sense of this analogy.

UNLOCKING MORAL KNOWLEDGE

To set up the theory behind the linguistic analogy, let me sketch three toy models of the sources of moral judgment. The first stems from the British Empiricists, and especially David Hume (1739/1978; 1748), by placing a strong emphasis on the causal power of emotions to fuel our moral judgments. As the top row of Figure 1 reveals, on this model, the perception of an event triggers an emotion which in turn triggers, unconsciously, an intuition that the relevant action is morally right or wrong, permissible or forbidden. I call an agent with such emotionally fueled, intuitive judgments, a *Humean creature*—illustrated by the character holding his heart; this is the model that today has been most eloquently articulated and defended by Antonio Damasio (1994, 2000, 2003) and Jonathan Haidt (2001).

The second model, illustrated in the middle row of Figure 1, combines the Humean creature with a *Kantian creature* (character scratching his brain), an agent who consciously and rationally explores a set of explicit principles to derive a moral judgment.¹ On this model, the perception of an event triggers, in parallel, both an emotional response as well as a conscious deliberation over principles. Sometimes these two distinct processes will converge on the same moral judgment, and sometimes they will diverge; if the latter arises, and conflict ensues, then some process must adjudicate in order to generate a final decision. This blended model has been defended most recently by Greene *et al.* (2001, 2004).

The third model, illustrated in the last row of Figure 1, captures the *Rawlsian creature* (Hauser, 2006; Hauser, Young *et al.*, in press b). On this model, event perception triggers an analysis of the causal and intentional properties underlying the relevant actions and their consequences. This analysis triggers, in turn, a moral judgment that will, most likely, trigger the systems of emotion and conscious reasoning. The single most important difference between the Rawlsian model and the other two is that emotions and conscious reasoning follow from the moral judgment as opposed to being causally responsible for them.

Though there are several other ways to configure the process from event perception to moral judgment, the important point here is that thinking about the various components in terms of their temporal and causal roles sets up the empirical landscape, and in particular, the role that neurobiological investigations might play in understanding these processes. Consider, for example, a strong version of the Humean creature, and the causal

¹ As I note in greater detail in Hauser (2006), Kant had far more nuanced views about the source of our moral judgments. In particular, he acknowledged the role of intuition, and classically argued that we can conform to moral law without any accompanying emotion. I use Kant here to reflect a rationalist position, one based on deliberation from clearly expressed principles.

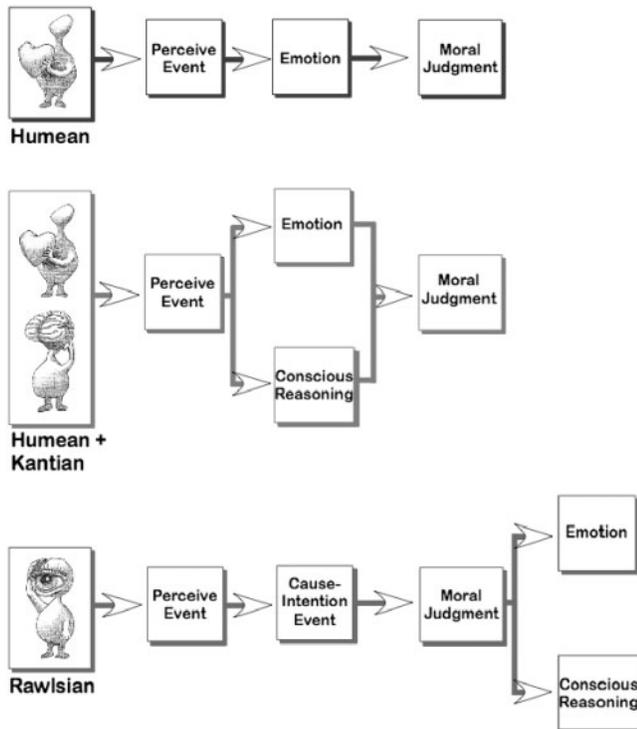


Fig. 1 Three toy models of the sources of our moral judgements.

necessity of emotion in fueling our moral intuitions. If we had perfect knowledge about the circuitry involved in emotional processing, and located a patient with damage to this circuitry, we would expect to find an individual who was incapable of delivering ‘normal’ moral judgments, where normal is defined in terms of both non-brain-damaged subjects as well as control patients with damage to non-emotion-relevant areas. In contrast, if the blended Humean-Kantian creature represents a better model, then damage to the emotional circuitry would only perturb those aspects of a moral dilemma that rely on the emotions; those aspects that are linked to the cool, rational components of the dilemma, such as the utilitarian payoffs, would be processed in an entirely normal fashion. Finally, if the Rawlsian creature experiences complete damage to the emotional circuitry, moral judgments will be indistinguishable from normals, but moral actions will be clearly distinguishable. On this model, emotions fuel actions or behaviors, be they approach or avoid, but play no role in mediating our judgments because judgments are guided by a system of unconscious knowledge. And if this model is correct, then psychopaths, when properly tested, will exhibit intact moral knowledge, but deficits with regard to normal behavior. Instead of inhibiting a desire to hurt or harm someone else, the lack of an emotional brake will lead to harming while also being fully cognizant of its moral impermissibility.

The discussion thus far cuts through only a small part of the moral territory, and focuses on the role of the emotions. There are other psychological processes that enter into the discussion, and are orthogonal to the concern of when emotions play a role. In particular, the Rawlsian creature focuses our attention on folk psychological processes including our ability to assess another’s intentions and desires, the perception of cause and effect, the calculation of utility and so forth.² Thus, for example, in both moral philosophy and law, a great deal of attention has focused on issues of responsibility, intent, negligence, foresight and desire. These are all psychological constructs that arise in both moral and nonmoral contexts, have been explored with respect to their ontogenetic emergence, and in some cases, have been linked to particular neural structures as a result of neuroimaging studies of normal subjects and neuropsychological studies of patient populations. For example, Saxe and colleagues (Saxe *et al.*, 2004; Saxe and Wexler, 2005) have demonstrated with fMRI that the temporo-parietal junction plays a critical role in false belief attribution, and Humphreys and colleagues (Samson *et al.*, 2004) have provided supporting evidence from patient studies. What is yet unknown is whether this area is recruited in the same or different way when false beliefs are married to moral dilemmas. Similarly, given that individuals with autism or Aspergers show deficits with respect to the attribution of mental states, including beliefs and desires, to what extent does this deficit impact upon their ability to distinguish morally relevant dilemmas where the consequences are the same but the means are different?

HOLES IN THE MORAL ORGAN

To date, neuropsychological reports have provided some of our deepest insights into the causally necessary role that certain brain regions and circuits play in cognitive function. Among the best studied of these are the language aphasias that range from the rather general deficits of comprehension or production to the more selective problems including deficits in processing vowels as opposed to consonants, and the selective loss of one language with the complete sparing of the other in bilinguals. And although the depth of our understanding of the neurobiology of language far outpaces other cognitive functions, due in part to the sophistication of linguistic theories and empirical findings, there remain fundamental gaps between the principles articulated by linguists to account for various details of grammatical structure, semantic expression, and phonological representation, and the corresponding neurobiological mechanisms. It should, therefore, come as no surprise that our understanding of the neuropsychology

² On the account presented here, both Kantian and Rawlsian creatures attend to causal/intentional processes, but the Kantian consciously retrieves this information and uses it in justifying moral action whereas the Rawlsian does so on the basis of intuition.

of moral knowledge is relatively impoverished, with only a glimmer of understanding. But since optimism, tainted with a healthy dose of criticism is most likely to engender enthusiasm for the potential excitement on the horizon, I use the rest of this essay to showcase some of the pieces that are beginning to emerge from patient populations, targeting the particular questions raised in the previous section.

The three toy models presented in Figure 1 target problems involving the temporal and causal ordering of processes, as well as the particular processes themselves. At stake here, as in any putative domain of knowledge, is the extent to which the processes that support both the operation and acquisition of knowledge are specific to this domain as opposed to domain general, shared with other mind-internal systems. Thus, we want to understand both those processes that support our moral judgments as well as those that are specific to morality as a domain of knowledge. To clarify this issue, consider a recent study (Koenigs *et al.*, in review) targeting the role of emotions in moral judgments, and involving patients with adult onset, ventromedial prefrontal [VMPC] damage, that have been carefully studied by Damasio, Tranel, Adolphs and Bechara (Bechara *et al.*, 1994, 1997; Damasio, 1994, 2000, 2003; Tranel *et al.*, 2000). Prior work on these patients indicated a deficit in making both immediate and future-oriented decisions. One explanation of this deficit is that these patients lack the kind of emotional input into decision-making that non-brain damaged subjects experience. That is, for normal subjects, decision-making is intimately entwined with emotional experience. In the absence of emotional input, decision-making is rudderless. Given this diagnosis, our central question was: do emotions play a causally necessary role in generating moral judgments? More specifically, we sought evidence that would adjudicate between the general role of emotions in socially relevant decisions, and the more selective role that emotions might play in morally relevant decisions, including their potential role as either the source or the outcome of our moral judgments.

To address these issues, it was necessary to dissect the moral sphere into a set of socially relevant distinctions. In the same way that early work in developmental psychology sought a distinction between social conventions and moral rules (Turiel, 1998, 2005), we sought a further set of distinctions within the class of problems considered moral. In particular, we presented ventromedial prefrontal patients, together with brain damaged controls and

subjects lacking damage, with a suite of scenarios, and for each, asked ‘Would you X?’³ The first cut through these scenarios contrasted non-moral dilemmas with two classes of moral dilemmas—impersonal and personal. Nonmoral dilemmas included situations in which, for example, a time-saving action would potentially be offset by a significant financial cost. Impersonal and personal moral dilemmas included cases providing options to harm one person in order to save many; the critical distinguishing feature between impersonal and personal was that the latter required some kind of physical harm with the target individual or individuals, whereas the former did not. The classic trolley problem provides a simple case: on the impersonal version, a bystander can flip a switch that causes a runaway trolley to move away from five people on the track to a side track with one person; on the personal version, a bystander can push a heavy person in front of the trolley, killing him but saving the five ahead.

The three groups—both brain damaged populations and the non-brain damaged controls—showed the same patterns of response for both non-moral dilemmas as well as impersonal moral dilemmas. Where a difference emerged was in the context of personal moral dilemmas: the ventromedial prefrontal patients were significantly more likely to say that it was permissible to cause harm to save a greater number of others, resulting in a strongly utilitarian response—*independently of the means*, it is always preferable to maximize the overall outcome or utility. These results suggest that the deficit incurred by the ventromedial patients does not globally impact upon social dilemmas, and nor does it more selectively impact upon moral dilemmas. Rather, damage to this area appears to selectively impact upon their judgments of personal moral dilemmas.

What do we learn from this pattern of results, and especially the causal role of emotions? Given that impersonal moral dilemmas are emotionally salient, we can rule out the strong claim that emotions are causally necessary for all moral dilemmas. Instead, we are forced to conclude that emotions play a more selective role in a particular class of moral dilemmas, specifically, those involving personal harm. In separate ratings by non-brain damaged subjects, all personal dilemmas were classified as more emotional than all impersonal dilemmas. But we can go further.

Looking at the range of personal moral dilemmas revealed a further distinction: some dilemmas elicited convergent and rapidly delivered answers (*low conflict*), whereas others elicited highly divergent and slowly delivered answers (*high conflict*). Consistently, the ventromedial patients provided the same moral judgments as the other groups for the low conflict dilemmas, but significantly different judgments for the high conflict dilemmas. Again, the ventromedial patients showed highly utilitarian judgments when contrasted with the others.

³ In the psychological literature on moral judgments, there is a considerable variation among studies with respect to the terminology used to explore the processes guiding subjects’ responses to different dilemmas. Thus, some authors use neutral questions as in ‘Would you X?’ or ‘Is it appropriate to X?’ (Greene *et al.*, 2001; 2004), others more explicitly invoke the moral dimension as in ‘Is it morally permissible to X?’, including queries that tap the more complete moral space by asking about the psychological positioning of an action along a Likert scale that runs from forbidden through permissible to obligatory (Cushman *et al.*, in press; Hauser *et al.*, in press b; Mikhail, 2000). At present, it is unclear how much variation or noise these framing effects have on the general outcome of people’s moral judgments.

For example, on the low conflict case of a teenage girl who wants to smother her newborn baby, all groups agreed that this would not be permissible; in contrast, on a divergent case such as Sophie's choice where a mother must either allow one of her two children to be tested in experiments or she will lose both children, VMPC patients stated that the utilitarian outcome was permissible (i.e., allow the one child to be tested), whereas the two other groups stated that it was not. Intriguingly, the low vs high conflict cases appear to map (somewhat imperfectly) on to a further distinction, one between self- and other-serving situations. Whereas the pregnant teen entails a selfish decision, Sophie's choice involves a consideration of harms to others. The VMPC patients showed the same pattern of judgments on the self-serving cases as the controls, but showed the utilitarian response on most of the other-serving cases.

Two conclusions emerge from this set of studies. First, the role of emotions in moral judgments appears rather selective, targeting what might be considered true moral dilemmas: situations in which there are no clear adjudicating social norms for what is morally right or wrong, and where the context is intensely emotional. One interpretation of this result is that in the absence of normal emotional regulation, VMPC subjects fail to experience the classic conflict between the calculus that enables a utilitarian or consequential analysis and the system that targets deontological or nonconsequential rules or principles (Greene and Haidt, 2002; Greene *et al.*, 2004; Hauser, 2006). When emotional input evaporates, consequential reasoning surfaces, as if subjects were blind to the deontological or nonconsequential rules. Second, we can reject both a strong version of the Humean creature as well as a strong version of the Rawlsian creature. Emotions are not causally necessary for generating all moral judgments, and nor are they irrelevant to generating all moral judgments. Rather, for some moral dilemmas, such as those falling under the category of impersonal as well as personal/low conflict/self-serving, emotions appear to play little or no role. In contrast, for personal/high conflict/other-serving dilemmas, emotions appear to play a critically causal role. This conclusion must, however, be tempered by our rather limited understanding of the representational format and content of emotions, as well as their neural underpinnings. My conclusions rely entirely on the claim that the ventromedial prefrontal cortex is responsible for trafficking emotional experiences to decision-making processes. If it turns out that other neural circuits are critically involved in emotional processing, and these are intact, then the relatively normal pattern of responses on impersonal cases, as well as personal/low conflict/self-serving cases, is entirely expected. To further support the Rawlsian position, we would need to observe normal patterns of moral judgments on these cases following damage to the circuitry associated with emotional

processing. Alternatively, if we observe deficits in the patterns of moral judgments, then we will have provided support for the Humean creature, and its emphasis on the causal role of emotions in generating moral judgments.

Patient populations will also prove invaluable for a different aspect of the three toy models, specifically, the relative roles of conscious and unconscious processes. As stated, there are two contrasting accounts of our moral judgments, the first appealing to conscious, explicitly justified principles, the second, appealing to intuitive processes, mediated by emotions, a moral faculty that houses inaccessible principles or some combination of the two. Based on a large scale Internet study involving several thousand subjects, my students Fiery Cushman, Liane Young, and I have uncovered cases where individuals generate robust moral judgments in the absence of generating sufficient justifications (Cushman *et al.*, in press; Hauser, Cushman *et al.*, in press a). For example, most people state that it is permissible for a bystander to flip a switch to save five people but harm one, but it is forbidden to push and kill the heavy man to save five people. When asked to justify these cases, most people are incapable of providing a coherent answer, especially since the utilitarian outcome is held constant across both cases, and the deontologically relevant means involves killing, presumed to be forbidden. When cases like these, and several others, are explored in greater detail, a similar pattern emerges with some dilemmas yielding a clear dissociation between judgment and justification while other dilemmas show no dissociation at all. In all cases, there is an operative principle responsible for generating the judgment. For example, people consistently judged harms caused by intent as morally worse than the same harms caused by a foreseen action (*Intention principle*); they judged harms caused by action as worse than inaction (*Action principle*) and finally, they judged harms caused by contact as worse than the same harms caused by noncontact (*Contact principle*). When asked to justify these distinctions, however, most subjects provided the necessary justification for the Action principle, slightly more than half justified the Contact principle, and extremely few justified the Intention principle. What these results suggest is that the Action principle, and to a lesser extent, the Contact principle, are not only available to conscious reflection, but appear to play a role in the process that moves from event perception to moral judgment to moral justification. In contrast, the Intention principle, with its distinction between intended and foreseen consequences, appears to be inaccessible to conscious reflection. Consequently, when moral dilemmas tap this principle, subjects generate intuitive moral judgments, using unconscious processes to move from event perception to moral judgment; when subjects attempt to justify their judgments, they will either state that they do not have a coherent explanation,

relying as it were on a hunch, or they will provide an explanation that is insufficient, incompatible with previous claims, or based on unfounded assumptions that have been added in an attempt to handle their own uncertainty.

Neuropsychological studies can help illuminate this side of the problem as well, targeting patients with damage to areas involved in mental state attribution, the maintenance of information in short-term working memory, and the ability to parse events into actions and sub-goals, to name a few. The distinction between intended and foreseen consequences is not restricted to the moral domain, appearing in plenty of nonmoral contexts. It is a distinction that plays a critical role in moral judgments, but is not specific to the moral domain. What may be unique to morality is how the intended-foreseen distinction—presumably part of our folk psychology or theory of mind—interfaces with other systems to create morally specific judgments. Take, for example, recent philosophical discussions, initiated by Knobe (2003a, b), on the relationship between a person's moral status and the attribution of intentional behavior. In the classic case, a Chief Executive Officer (CEO) has the opportunity to implement a policy that will make his company millions of dollars. In one version of the story, implementing the policy will also *harm* the environment, whereas in the other version, it will *help* the environment. The CEO implements the policy, and in the first case the company makes millions but also harms the environment, whereas in the second case, they also make millions but help the environment. When Knobe asked subjects about the CEO's decision, they provided asymmetric evaluations: they indicated that the CEO intentionally harmed the environment in the first case, but did not intentionally help the environment in the second case. One prediction we might derive from this case, and others showing parallel asymmetries, is that moral status is intimately intertwined with our emotions, with harm generating greater attributions of blame relative to help. If this pattern relies on normal emotional processing, then the ventromedial patients should show a different pattern from normals. They do not (Young et al., in press). This suggests that emotions are not necessary for mediating between moral status and intentional attributions. Rather, what appears to be relevant is the system that handles intentionality, which appears (based on prior neuropsychological tests) intact in ventromedial patients.

MORAL METAPHORS

Is the moral faculty like the language faculty? Is the moral organ like the liver? At this stage, we do not have answers to either question. The analogy to language is a useful heuristic in that it focuses attention on a new class of questions. Lkening the moral organ to a liver is also useful, even if metaphorical. By thinking about the possibility of a moral organ, again in the context of language, we seek

evidence for both domain-general and specific processes. That is, we not only wish to uncover those processes that clearly support our moral judgments, but in addition, identify principles or mechanisms that are selectively involved in generating moral judgments. For example, though emotions arise in moral and nonmoral contexts, and so too do the mental state representations such as intentions, beliefs, desires and goals, what may be unique is the extent to which these systems interface with each other. What is unique to the moral domain is how the attribution of intentions and goals connects with emotions to create moral judgments of right and wrong, perhaps, especially when there are no adjudicating moral norms.

The way that I have framed the problem suggests that we consider our moral faculty as anatomists, dissecting the problem at two levels. Specifically, we must first make increasingly fine distinctions between social dilemmas, especially along the lines of those sketched in the last section—nonmoral *vs* moral, personal *vs* impersonal, self-serving *vs* other-serving. Within each of these categories, there are likely to be others, contrasts that will emerge once we better understand the principles underlying our mature state of moral knowledge. Second, we must use the analysis from part 1 to motivate part 2, that is, the selection of patient populations that will help illuminate the underlying causal structure. The contrast between psychopaths and ventromedial patients provides a critical test of the causal role of emotions in both moral judgments (our competence) and moral behavior (our performance). The prediction is that psychopaths will have normal moral knowledge, but defective moral behavior.⁴ Another breakdown concerns the shift from conventional rules to moral dilemmas. Nichols (2002, 2004) has argued that moral dilemmas emerge out of the marriage between strong emotions and normative theories. In one case, subjects judged that it was impermissible to spit in a wine glass even if the host said it was okay. Though this case breaks down into a set of social conventions, it is psychologically elevated to the status of a moral dilemma due to the fusion between norms and strong emotions.⁵ To nail this problem, it is necessary to explore what happens to subjects' judgments when some relevant piece of neural circuitry has been damaged. For example, patients with Huntington's chorea experience a fairly selective deficit for disgust, showing intact processing of other emotions (Sprenkelmeyer *et al.*, 1996, 1997). In collaboration with Sprenkelmeyer, Young and I have begun testing these patients. Preliminary evidence suggests that they are normal on the Nichols' cases, but abnormal on others. For example, when normal,

⁴ Though Blair has already demonstrated that psychopaths fail to make the conventional/moral distinction, suggesting some deficit in moral knowledge, this distinction is not sufficiently precise with respect to the underlying psychological competence and nor does it probe dilemmas where there are no clear adjudicating rules or norms to decide what is morally right or wrong.

⁵ To be clear, Nichols is not claiming that spitting in a glass is, necessarily, a moral dilemma. Rather, when a social convention unites with a strong emotion such as disgust, that transgressions of the convention are perceived in some of the same ways that we perceive transgressions of unambiguously moral cases.

control subjects read a story in which a man who had been married for 50 years decides to have intercourse with his now dead wife, they judged this to be forbidden; in contrast, patients with Huntington's said the opposite: intercourse is perfectly permissible especially if the husband still loves his wife.

There is something profoundly interesting about our unquestioned willingness to swap an unhealthy liver for anyone else's healthy liver, but our unquestioned resistance to swap brain parts. One reason for this asymmetry is that our brain parts largely determine who we are, what we like, and the moral choices we make. Our livers merely support these functions, and any healthy liver is up to the job. But the radical implication of situating our moral psychology inside a moral organ is that any healthy moral organ is up to the job. What the moral organ provides is a universal toolkit for building particular moral systems.

REFERENCES

- Bechara, A., Damasio, A., Damasio, H., Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50, 7–15.
- Bechara, A., Damasio, H., Tranel, A., Damasio, A. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, 275, 1293–5.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York, NY: Praeger.
- Cushman, F., Young, L., Hauser, M.D. (in press). The role of conscious reasoning and intuition in moral judgments: testing three principles of harm. *Psychological Science*.
- Damasio, A. (1994). *Descartes' Error*. Boston, MA: Norton.
- Damasio, A. (2000). *The Feeling of What Happens*. New York: Basic Books.
- Damasio, A. (2003). *Looking For Spinoza*. New York: Harcourt Brace.
- Dwyer, S. (1999). Moral competence. In: Maurasugi, K., Stainton, R. editors. *Philosophy and Linguistics*. Boulder: Westview Press, pp. 169–90.
- Dwyer, S. (2004). How good is the linguistic analogy, [WWW]. Available: www.umbc.edu/philosophy/dwyer [2004, February 25, 2004].
- Greene, J. D., and Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science*, 6, 517–23.
- Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400.
- Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–8.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–34.
- Harman, G. (1999). Moral philosophy and linguistics. In: Brinkmann, K. editor. *Proceedings of the 20th World Congress of Philosophy*, Vol 1, Ethics: Bowling Green: Philosophy Documentation Center, pp. 107–15.
- Hauser, M.D. (2006). *Moral Minds: How Nature Designed Our Sense of Right and Wrong*. New York: Ecco/Harper Collins.
- Hauser, M.D., Cushman, F., Young, L., Jin, R.K.-X. (in press a). A dissociation between moral judgments and justifications. *Mind and Language*.
- Hauser, M.D., Young, L., Cushman, F. (in press b). Reviving Rawls' Linguistic Analogy: Operative principles and the causal structure of moral actions. *Moral Psychology*, Vol 2, Cambridge: MIT Press.
- Hume, D. (1739/1978). *A Treatise of Human Nature*. Oxford: Oxford University Press.
- Hume, D. editor. (1748). *Enquiry Concerning the Principles of Morality*. Oxford: Oxford University Press.
- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, 63, 190–3.
- Knobe, J. (2003b). Intentional action in folk psychology: an experimental investigation. *Philosophical Psychology*, 16, 309–24.
- Koenigs, M., Young, L., Adolphs, R., et al. (in review). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*.
- Mikhail, J.M. (2000). Rawls' linguistic analogy: A study of the 'generative grammar' model of moral theory described by John Rawls in 'A theory of justice'. Unpublished PhD, Cornell University, Ithaca.
- Mikhail, J.M. (in press). *Rawls' Linguistic Analogy*. New York: Cambridge University Press.
- Nichols, S. (2002). Norms with feeling: toward a psychological account of moral judgment. *Cognition*, 84, 221–36.
- Nichols, S. (2004). *Sentimental Rules*. New York: Oxford University Press.
- Rawls, J. (1951). Outline of a decision procedure for ethics. *Philosophical Review*, 60, 177–97.
- Rawls, J. (1963). The sense of justice. *Philosophical Review*, 72, 281–305.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.
- Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7, 499–500.
- Saxe, R., Carey, S., Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55, 1–27.
- Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporoparietal junction. *Neuropsychologia*, 43, 1391–9.
- Sprenghelmeyer, R., Young, A.W., Calder, A.J., et al. (1996). Loss of disgust: perception of faces and emotions in Huntington's disease. *Brain*, 119, 1647–65.
- Sprenghelmeyer, R., Young, A.W., Pundt, I., et al. (1997). Disgust implicated in obsessive-compulsive disorder. *Proceedings of the Royal Society of London B*, B264, 1767–73.
- Tranel, D., Bechara, A., Damasio, A. (2000). Decision making and the somatic marker hypothesis. In: Gazzaniga M. editor. *The New Cognitive Neurosciences*. Cambridge: MIT Press, pp. 1047–61.
- Turiel, E. (1998). The development of morality. In: Damon, W. editor. *Handbook of Child Psychology*. New York: Wiley Press, pp. 863–932.
- Turiel, E. (2005). Thought, emotions, and social interactional processes in moral development. In: Killen, M., Smetana, J. G. editors. *Handbook of Moral Development*. Mahwah: Lawrence Erlbaum Publishers.
- Young, L., Cushman, F., Adolphs, R., Tranel, D., Hauser, M.D. (in press). Does emotion mediate the relationship between an action's moral status and its intentional status? Neuropsychological evidence. *Cognition and Culture*.