

AUTHOR INTERVIEW

The Bookshelf talks with Marc Hauser

[Greg Ross](#)

Oscar Wilde said, "Morality, like art, means drawing a line someplace." But how do we learn where to draw these lines? It's commonly understood that moral rules are instilled in church, school and home, but Harvard psychologist Marc Hauser believes that they have a deeper source—an unconscious, built-in "moral grammar" that drives our judgments of right and wrong.

Widely known for his studies of animal cognition (see "[What Do Animals Think About Numbers?](#)" in the March-April 2000 *American Scientist*), Hauser has long been intrigued by the nature of human moral judgment (interested readers can take his Web-based [Moral Sense Test](#)). He says the human sense of right and wrong, which evolved over millions of years, precedes our conscious judgments and emotions, providing a hidden engine of moral intuition that's shared by people around the world. "Our moral instincts are immune to the explicitly articulated commandments handed down by religions and governments," he writes. "Sometimes our moral intuitions will converge with those that culture spells out, and sometimes they will diverge." In *Moral Minds* (Ecco, available August 22) Hauser draws ideas from the social and natural sciences, philosophy and the law to support his own findings for an unconscious moral instinct.

[click for full image and caption](#)

American Scientist Online managing editor Greg Ross interviewed Hauser by e-mail in July 2006.

Can you describe what you mean by a moral grammar?

The core idea is derived from the work in generative grammar that [MIT linguist Noam] Chomsky initiated in the 1950s and that the political philosopher John Rawls brought to life in a short section of his major treatise *A Theory of Justice* in 1971. In brief, I argue that we are endowed with a moral faculty that delivers judgments of right and wrong based on unconsciously operative and inaccessible principles of action. The theory posits a universal moral grammar, built into the brains of all humans. The grammar is a set of principles that operate on the basis of the causes and consequences of action. Thus, in the same way that we are endowed with a language faculty that consists of a universal toolkit for building possible languages, we are also endowed with a moral faculty that consists of a universal toolkit for building possible moral systems.

By grammar I simply mean a set of principles or computations for generating judgments of right and wrong. These principles are unconscious and inaccessible. What I mean by unconscious is different from the Freudian unconscious. It is not only that we make moral judgments intuitively, and without consciously reflecting upon the principles, but that even if we tried to uncover those principles we wouldn't be able to, as they are tucked away in the mind's library of knowledge. Access comes from deep, scholarly investigation.

Paralleling language, the notion of grammar that has been developed in modern linguistics is virtually incomprehensible outside the field of linguistics. The grammar we learned in school has virtually no resemblance to the grammatical principles uncovered

by linguists. In the same way, once we delve deeper into our moral faculty we will uncover principles that are virtually unrecognizable from the social norms that we articulate and live by in our day-to-day lives. And in the same way that the unconscious but operative principles of language do not dictate the specific content of what we say, if we say anything, the unconscious but operative principles of morality do not dictate the specific content of our moral judgments, nor whether we in fact choose to help or harm others in any given situation.

If there is an innate moral instinct, why was it selected? What advantages does it bring?

First, the only way for selection to work, if we are discussing biological as opposed to cultural evolution, is for the trait to have some genetic, heritable component that is variable. Now the challenge comes in working out the selective advantage of such a capacity, and here there are at least two options.

First, it is possible that the moral instinct was originally selected due to its fitness consequences for maintaining social norms, some of which may have evolved long before humans emerged on earth. That is, the moral faculty provides a set of principles for cooperating, for punishing cheaters, for determining the conditions in which helping is obligatory, and so forth.

Second, it is possible that some of the computations that underlie our moral instinct evolved for reasons that are not specific to morality but were subsequently co-opted or adopted for morality, and then subject to a round of selection. For example, take the fact that many moral decisions are based upon whether an action was intended or accidental. If someone harms another, it is essential to assess whether the harm was intended or the result of an accident. If accidental, was it due to negligence? Though the intended/accidental distinction is critical to our moral evaluations, the ability to distinguish these two causal factors appears in non-moral situations: Though I am a fairly good tennis player, sometimes I hit a winner because I really aimed for a particular spot inside the line, and sometimes I accidentally hit the spot. The consequence is the same: I hit a winner.

Early in human development, children appear sensitive to these hidden psychological causes, appreciating that not all consequences are equal. They appreciate that the means of achieving a particular consequence matter for both moral and non-moral situations.

In this light, how should we regard social institutions such as education, religion and law, which transmit moral codes?

We are only at the beginning stages of this work, and the theoretical implications are fairly radical. But if the notion of a universal moral grammar is right, then it raises some interesting issues with respect to the role of experience, coming in from parents, teachers, religious institutions and so on.

Like language, the notion of a universal moral grammar should not be equated with the rejection of cultural variation. Like language, cross-cultural variation is expected. But the moral faculty will place constraints on the range of cross-cultural variation and thus limit the extent to which religion, law or teachers can modify our intuitive moral judgments.

For example, in a large sample of moral dilemmas that involve questions concerning the permissibility of harming other individuals, we have found no significant differences in the pattern of moral judgments between people who are religious and people who are

atheists. Similarly, for a certain class of dilemmas we have found little effect of education. This is not to say that education and religion have no impact on our moral psychology. Rather, it is to say that certain aspects of our moral intuitions seem to be immune to such experience.

Where, then, does experience play a role? Here it is important to make a distinction between how we judge particular moral dilemmas and what we actually do. If I ask whether it is permissible for you to intentionally kill someone else, your initial response might be "No!" But upon reflection, you will soon realize that there are several contexts in which it is permissible to intentionally kill someone. For example, it is permissible in situations of self-defense and in war. And in some cultures or societies, it is permissible to have an abortion, to commit infanticide and to kill an unfaithful spouse. What this reveals is that associated with universal principles of harming and helping others are parameters or switches that experience flips, thereby creating small pockets of variation. This parallels language, where there is parametric variation for components of language, such as the order of subject, object and verb.

What remains to be explored is the extent to which experience might have effects on what we do, as opposed to how we judge certain dilemmas. Take, for example, the observation, confirmed by many studies, that people judge actions that cause harm as worse than omissions that result in the same harm. This bias pops up in all sorts of biomedical cases, of which the distinction between active and passive euthanasia is one prominent example. Thus, most countries allow for passive euthanasia, in which life support is terminated, but block actively ending someone's life by means of drug administration. Though the intent is the same in both cases—end the suffering of a patient with a terminal illness—and the consequence is the same—the patient dies—many members of the medical community see a need to make a legal distinction. What I argue, however, is that if people can be made aware of this bias, they may be able to end up at a different end point, one that is far more consistent with doctors and nurses working in the trenches who see the distinction as meaningless; and for some, there is the strong sense that passive euthanasia is less humane than active euthanasia, given that the patient will often suffer for a longer period of time.

If this faculty were lacking, in an individual or in our species, what effect would we see?

It's hard to say what the answer to this is at this point, because we have only begun to flesh out the theory and run the relevant experiments. That said, we have also begun to test patients with damage to particular parts of the brain, and thus our understanding is increasing at a rapid clip, with exciting results on the horizon. Let me give a few illustrative examples.

One of the challenging implications of the idea that our moral faculty is home to a universal moral grammar is that we generate intuitions about which actions are morally right or wrong prior to generating any emotions. On this view, emotions follow from our moral judgments, as opposed to preceding them. And on this view, emotions guide what we do as opposed to how we judge particular moral dilemmas. Following this through, we would predict that the clinical problem observed among psychopaths comes not from damage to their moral faculty, but rather from damage to the systems of emotion that lead from perception to action.

To explore this idea, we are testing psychopaths in collaboration with James Blair of the NIH. The prediction is that psychopaths will show normal patterns of responses to various moral dilemmas, but show deficits in what they actually do. That is, they will have intact

moral knowledge, but deficits in morally relevant actions due to problems of emotional control.

Though we have yet to collect the relevant data on psychopaths to affirm or reject this idea, we have collected data on a closely related population of patients with damage to the frontal lobes. These patients have been characterized as "acquired sociopaths," given that the damage leads to socially inappropriate behavior. The most famous of these patients is Phineas Gage, who in the 19th century suffered severe damage to the frontal lobes due to injury from a railroad tamping iron and went from a model citizen to an individual who lost his job, lost all sense of social appropriateness and ultimately became a vagrant, aimlessly wandering from town to town.

In our recent studies, collaborating with a patient population that has been carefully studied by the neuroscientist Antonio Damasio and his colleagues, we have found an exciting and highly selective deficit. Whereas these patients show normal patterns of responses to a relatively large class of moral dilemmas, they show highly abnormal responses on one specific type of dilemma. In particular, where the action involves personal contact with another individual, and where the choice is between harming one versus many, and there are no clear social norms available to decide, these patients consistently take the utilitarian route, selecting the option that yields the greatest good regardless of the means required to achieve such ends. Thus, damage to this particular area of the brain, one that connects emotional processing with high-level decision making, yields a highly selective deficit in moral judgment. Of course, if you are a utilitarian, your interpretation will be different! You will think that it is because of irrational emotions that we don't all think like utilitarians, seeing the overall good as the only relevant moral yardstick.

You've studied the differences between human and animal minds. Do you believe that other species have moral instincts?

What I believe we can say at present is that animals have some of the key components that enter into our moral faculty. That is, they have some of the building blocks that make moral judgments possible in humans. What is missing, with the strong caveat that no one has really looked, is evidence that animals make moral judgments of others, assigning functional labels such as "right," "wrong," "good," "bad" and so on to either actions or individuals.

In many ways, our understanding of animals is not even ripe for the picking, because almost all of the work that is relevant to morality entails studies of what animals *do* as opposed to how they judge what others do. Thus, we have beautiful accounts of how animals behave during cooperation and competition, including observations of how individuals respond to personal transgression, such as taking food in the presence of a more dominant animal. But what is missing are observations and experiments that systematically address what counts as a transgression or expectation for helping or harming, when the observer is not directly involved. In the same way that we can judge an act as gratuitously violent even when it doesn't concern us directly, we want to understand how animals perceive violations of social norms, including what they expect and what they consider anomalous.

These caveats aside, we are beginning to understand some of the relevant building blocks that are not specific to morality but play a key role. For example, in work on tamarin monkeys and chimpanzees, there is evidence that individuals distinguish between intentional and accidental actions. This is important because it shows, contrary to many prior claims, that animals are attending to more than the consequences. If

animals lacked this capacity, then they wouldn't even be in the running for consideration as moral agents. Further, animals seem to distinguish between animate and inanimate objects, which, again, is not a specifically moral distinction but is critically involved in moral judgments. Gratuitously smacking a candy machine may be perceived as odd but has no moral weight; gratuitously smacking a baby is not only odd but morally wrong!

What implications does a moral instinct have for current ethical debates like euthanasia and stem-cell research?

As I briefly mentioned above, the more we understand about our moral instinct, the more we may be able to make people aware of some of the psychological biases that they carry forward in their moral deliberations. Part of the work will come from understanding which aspects of our moral instinct tap specifically moral psychological processes and which tap more general systems of the mind. Thus, ethical questions such as euthanasia and stem-cell research entail questions about personhood, ownership, actions versus omissions, and responsibility; none of these are specifically moral issues. Thus, the fact that many perceive active euthanasia as worse than passive euthanasia may stem from a non-moral bias to see actions as more intentional than omissions. For example, if I intentionally knock off my mother's priceless vase from the table, that is worse than if I fail to catch it from falling off the table but could have. In the end, my mother's vase is broken, but most will see my knocking it off the table as worse.

To be explicit, the theory that I have developed in *Moral Minds* is a descriptive theory of morality. It describes the unconscious and inaccessible principles that are operative in our moral judgments. It does not provide an account of what people *ought* to do. It is not, therefore, a prescriptive theory of morality. That said, I am certain that a better understanding of the descriptive principles will ultimately shape how we develop our prescriptive theories, be they legal or religious. Here's how and why. Theories or statements concerning what we should do are based on notions about the human condition, about a life well lived and the conditions that support it. We think about freedom and justice, and we then explore this space, constantly reflecting upon the current situation and whether things could be better. But ultimately, any change that we attempt to impose because we think things ought to be different will potentially be opposed or resisted by our evolved psychology. Thus, though our biology does not dictate what we ought to do, we are much more likely to implement changes in our legal, religious or political policies by attending to the psychological predispositions that our biology handed down, and that local culture and recent history may have tuned up.