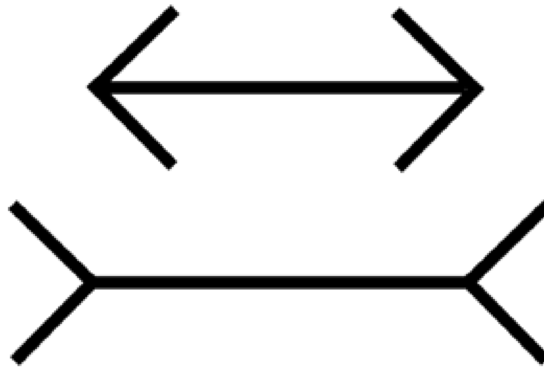
science & spirit
exploring things that matter
(Spring 2006)

Could a covert moral grammar underlie the illusion of conscious reasoning?

Marc D. Hauser



Stare at this illustration. Is the horizontal line on top shorter, longer, or equal to the horizontal line below? Though you, like everyone else, presumably see the horizontal line on top as shorter than the one below, the two lines are identical. Now that I have blown the cover on this illusion, try to convince yourself that the two lines are equal. Try convincing your visual system. Any luck? I didn't think so.

Visual illusions, like the Muller-Lyer illustration above, are interesting because no matter how long we stare at them, and no matter how well we know the trick, we can't convince those visual cortices of ours to see things differently. I would be surprised if anyone found this discovery upsetting. If anything, we should be intrigued, puzzled by our susceptibility to visual illusions and our inability to exert the power of free will to alter what we perceive. But what if I insisted that some aspects of our moral judgments are also illusory? What if I told you that what you perceive as the conscious and rational process leading from clearly articulated principles to confident moral judgments is often an illusion? What if I suggested that operating covertly, beneath the radar of our awareness, is a moral faculty, an engine for generating intuitions about forbidden and permissible actions? I assume that many people would find this claim upsetting, possibly absurd, and even derogatory.

The claim that we have been duped by an illusion does not imply that our moral judgments are false or necessarily misguided. Rather, the claim is that we often deliver our moral verdicts as rapidly and unconsciously as we blink an eye in response to a puff of air or generate judgments concerning the grammaticality of a sentence, the beauty of a painting, or the humor in a joke. Still, this runs

counter to our strong, albeit illusory, belief that we deliver our moral verdicts based on carefully considered reasons or principles, moving rationally among the possible alternative solutions to a dilemma and then alighting on one using utilitarian arguments that maximize overall practicality or deontological reasons that focus on actions that are, by definition, right or wrong. For example, a deontological stance based on rules of conduct would argue that killing is wrong and that it is therefore impermissible to take a healthy person's life in order to save five people in critical care. A utilitarian, meanwhile, would claim that the means are irrelevant and that if five can be saved, the one person's life is worth taking. When thousands of subjects participated in our Internet-based morality survey, they—utilitarians and deontologists alike—were confronted with a set of dilemmas that challenged the consistency of their moral convictions. Ninety-eight percent of our subjects said it was impermissible for a doctor to take the life of a healthy person in order to use his organs to save five patients in critical condition. In contrast, ninety percent of these subjects said it was permissible for an observer to redirect a train onto a side track, causing one person to die but saving five others who had been in harm's way. Utilitarians should perceive both cases as permissible. Deontologists should see both cases as forbidden. But the percentages show that neither philosophical perspective can do justice to these paired cases. No matter what kind of moral gymnastics our subjects try, they will ultimately be left with an unsatisfactory conclusion—an insufficient justification for their clear and universally shared moral judgments.

Driving our moral judgments is a universal moral grammar, a faculty of the mind that evolved over millions of years to include a set of principles for building a range of possible moral systems. As with language, the principles that make up our moral grammar fly beneath the radar of our awareness. As universal, biologically endowed principles of the mind, they are safeguarded against our consciously held beliefs, be they handed down from university classes, legal scholarship, mom and dad, political propaganda, or religious doctrine. Where culture enters the story is in setting particular parameters, early in development, as each child builds her signature moral system. Evidence for this position is beginning to emerge, but enough is already in play to foreshadow a revolution in our understanding of morality and its origins.

Extracting our moral grammar

If we want to understand how people respond to certain moral dilemmas, it is of little use to present situations in which the answers or responses have been well practiced. If I ask you to decide whether abortion, euthanasia, or embryonic stem cell research is morally right or wrong, your response probably won't give me much insight into the nature of your moral psychology because it's very likely that you've already formed an opinion, one that's based on frequent rehearsals of the argument and armored with robust emotional defenses. The psychology that enters into your decision has shifted from something that may once have been instinctive and unconscious to something that is conscious, rehearsed like a mantra, and sculpted by the local culture. To get at our intuitive psychology, philosophers and psychologists interested in the sources of our moral judgments have, especially in recent years, shifted the line of investigation, focusing instead on how people respond to unfamiliar, often highly artificial, sometimes absurd scenarios. The advantage of presenting fictional cases is that there can be no prior judgment or opinion, and carefully crafted dilemmas will still capture some of the essential psychological ingredients that presumably enter into our judgments about real-world cases such as euthanasia and abortion.

Consider the following example from philosopher James Rachels: A greedy uncle intends to snuff out his nephew in order to collect his inheritance. In one version of the story, the uncle enters the bathroom and drowns his nephew. In a second version of the story, the uncle enters the bathroom with murder on his mind, finds his nephew already drowning in the bathtub, and walks out, leaving the nephew to die. In both cases, the uncle intends to kill his nephew and in both cases, the nephew dies. But in the first scenario, the uncle acts to kill the nephew whereas in the second scenario, the uncle omits an act, thereby allowing the nephew to die. If you were on a jury deciding the uncle's fate, would you try him differently for these two cases? Many people initially feel that actions causing harm are worse than omissions causing the same

harm. Further reflection, however, leads everyone to see these as indistinguishable, and legal systems agree. The intent was malicious in each case, and the consequences identical. In one case, the uncle acts to end the nephew's life. In the other, the uncle omits the act of saving, which is not legally permissible because there were no costs associated with saving the nephew from drowning. We shouldn't see a difference between the psychology of action and omission, and, at least in this case, we don't.

In most countries, legal systems block active euthanasia (giving a terminal patient an overdose—so-called "mercy killing") while permitting passive euthanasia (allowing the patient to die by ending life support). Upon reflection, many see it as valid to support passive euthanasia, but block active euthanasia. But if we unpack this case as we did for the greedy uncle, we uncover an identical set of psychological ingredients: In both cases, the doctor's intent is to end the patient's pain and suffering, and in the end, the patient dies. Though our instincts may pull us to see action as worse than omission, on a practical level we should focus people's attention on the parameters that matter in these cases. Interestingly, doctors and nurses working in the trenches of real-world cases see the distinction between active and passive as patently absurd, often risking their careers by carrying out illegal acts of euthanasia—for completely humane reasons. Some countries, such as The Netherlands and Belgium, have even banished the distinction between active and passive euthanasia, leaving the issue up to a board of doctors and the patient's family.

So the fictional cases that help capture the nature of our intuitions also help uncover the psychological ingredients that enter into our moral judgments. These scenarios allow us to reveal the principles and parameters, analogous perhaps to those observed in language, which constitute our moral grammar. Some of the psychological ingredients we discover may have ancient origins and early developmental expressions—having evolved for reasons that have little or nothing to do with morality per se, but now playing a role in our moral psychology—including biasing effects that may be hard to discard. As research by Jonathan Baron and my own lab shows, people instinctively see actions as worse than omissions even when the consequences are the same. This bias may emerge, however, because we, and all other animals, are more likely to perceive the causal connection between actions and consequences than between omissions and consequences. When a man hits and injures another, his intentions are clear. When a man fails to catch another who is injured by a fall, his intentions for the omitted action are unclear. This simple and highly general learning account is reasonable, with supporting evidence showing up in the animal kingdom in the context of deception: Animals, of a variety of shapes, colors, and sizes, are far more likely to deceive by committing acts of omission than commission. For example, rhesus monkeys give characteristic calls when they find food, announcing their discovery to close group affiliates, often kin. But sometimes they omit these calls, attempting to silently steal away with the spoils of the day. If they succeed, score! If they are caught red handed by their group mates, they are beaten.

What these conclusions lead to is an exciting new research program that unites the sciences of the mind with age-old philosophical questions about the origins of moral knowledge. And this is precisely what my students and I have embarked upon, summarized in my forthcoming book *Moral Minds: how nature designed our universal sense of right and wrong.*

Universality and cultural immunity

The moral grammar perspective, like its twin in linguistics, makes several explicit predictions. If there is a universal moral grammar, then there should be universally held principles. If these principles operate like those underlying language, then we will be unconscious of their operation in generating moral judgments. In this sense, "unconscious" is nothing like what Sigmund Freud espoused, in which the principles were once part of our conscious life, but subsequently suppressed because they reminded us of some traumatic childhood experience. Rather, the principles are not only unconscious, but also inaccessible. Recovery comes only from intense scholarship, from studying hundreds of moral dilemmas and attempting to uncover principles that account for the patterns of judgments. Thus, in the same way that it took years of scrutiny by linguists to understand why it is OK to say "he

loves her" but not "her loves he," it has taken years of scrutiny by moral philosophers, and now scientists, to discover why we intuitively see intended harms as worse than harms that are foreseeable but not prevented. And from a perspective that sees the universal principles as unconsciously operative, we derive a further prediction: When we ask people to justify their moral judgments, they should come up empty-handed as they have no prior access to the process leading up to judgment. Finally, if the principles are innate, universal, and unconsciously operative, then we should find consistency across both demographic and cultural variation, and in some cases, an immunity to the influences of education handed down from government and religion. In the same way that we know we are being duped by a visual illusion, but can't convince our visual cortices to see it our way, religious and legal doctrine may drill into our heads the wrongness of killing, but have little impact on the cases in which we intuitively think it is permissible to harm another.
Marc: I really want a new graf here to make sure the following thought stands alone and does not get lost at the end of a long graf.Ok.

To be clear, I am neither suggesting a rejection of cross-cultural variation nor implying that we are born with a code of conduct that completely dictates what we do. Rather, I advocate the position that we are born with a universal moral grammar that equips us with a set of tools for building different kinds of moral systems, perhaps similar in kind to our language faculty's capacity to build different languages; these tools are principles concerning the nature of action, its causes and consequences. Cultural variation emerges when a culture sets one or more parameters in a different way. But sometimes, a culture may find that it hits a wall of resistance from the universal moral grammar that pushes us to see things differently. And in this case, the best bet is for scholars to make our evolved biases transparent, writing them on yellow sticky notes as reminders of how our minds evolved intuitions that may have once been adaptive, but today are no longer so..

My colleagues and I have addressed these questions by testing a wide variety of populations, including adults in different countries, with different demographic and cultural backgrounds; children prior to the age at which school might impact their moral attitudes; patients with selective damage to different parts of the brain believed to be relevant to moral analysis; and studies of animals that may have access to some of the central psychological distinctions involved in generating a moral verdict. In a large-scale study of several thousand subjects from more than 100 countries, presenting moral dilemmas focused on the nature of harm and, specifically, the permissibility of killing another individual, we have found three striking results. First, the effects of demographic and cultural variables on the pattern of moral judgments are insignificant. Second, there is a dissociation between judgment and justification, such that people rarely produce coherent justifications for their moral verdicts of right and wrong. And third, there are three principles that appear to unconsciously guide people's judgments, when consequences are held constant: People judge intended harms as worse than foreseeable harms, harms resulting from action as worse than harms resulting from omission, and harms involving physical contact as worse than those involving no contact.

These judgments appear to be immune to educational and religious background. Controlling for age, people with only a high school education are no different than those with advanced degrees when it comes to judging the permissibility of harming another in certain contexts. Similarly, people with strong religious backgrounds are no different than atheists and agnostics. Admittedly, we have not sampled every culture, and importantly, most of the evidence comes from people with Internet access; this population most likely has greater international awareness than most and, by definition, is affluent enough to put in phone lines and cables. That said, two pieces of evidence make these results striking and provide further support for the idea that some aspects of our moral psychology are immune to cultural background. First, as we have seen, straightforward utilitarian or deontological perspectives fail to provide a functional handbook for navigating these dilemmas. In a classic pair of problems from moral philosophy, a person can flip a switch to prevent a trolley from killing five others, diverting it onto a side track where it will kill one; or instead, the person can push a man onto the tracks, killing him, but saving the five ahead. The utilitarian option, favored by some religions and cultures,

would have subjects always pick saving many. The deontological option, favored by others, would have subjects avoid killing. But these biases don't work consistently across cases, leaving subjects in a further quandary as their judgments get pushed around. What starts out looking like a rational position, backed by religious and legal doctrine or cultural norms, ends up as inconsistent, irrational, wishy-washiness. Second, some subjects are clearly religious while others are clearly not, and yet their judgments are no different in many of these cases—and neither are their justifications; independently of religious background, if an atheist or agnostic provides an incoherent explanation for his judgment, so too does a Jew, Catholic, Muslim or Buddhist.

Some may wish to argue that evidence of universality is no more evidence for a biologically evolved faculty than it is evidence for a God or divine power that universally endowed all humans with these views. What is universal is simply the insignia of divine creation, they would say. But let's unpack the logic into the observations, inferences, and conclusions. What I have presented here, and in our scientific publications, is evidence that there are no statistically significant differences in the patterns of judgments between people with or without religious backgrounds for a host of moral dilemmas involving harming and helping others. When people with religious backgrounds judge these cases, religious doctrine does not provide a set of bulletproof principles for adjudicating between the cases. We can interpret this result in two ways: either a divine power created our universal moral sense or evolution did. At this point, we reach an empirical impasse because there is no proof for or against divine power. But for those who think that divine power created our universal moral sense, there is a problem: How do we explain the observation of universal intuitions regarding harming and helping others and the fact that there are some principles held by some religions that are not universal? If you are religious, and believe that your religion, with its set of doctrinal principles is perfectly aligned with a divine power's principles, then you have to agree that universality is derived from some other source than the divine. Biology would be the logical candidate. Alternatively, you could argue that a divine power *is* the source of the universal, but religions choose to use other principles. But if religions are free to choose in this way, deriving their inspiration from something other than the divine, then much of the motivation and emotion underlying formal religion is in jeopardy.

The moral grammar perspective is an attempt to uncover some of the unconscious principles mediating our moral judgments and the ways in which cultural factors can push around the nature of these judgments. In some cases, we will uncover the impact of religious and political beliefs, and in others, we will find universality and immunity. As I have stated, the linguistic analogy does not deny cultural variation. Rather, it predicts certain universals concerning the permissibility of harming and helping others, while providing constraints on the range of cultural variation. Thus, in no culture is killing always wrong or helping always obligatory. Each culture has principles that dictate when killing is permissible and when helping is forbidden. The linguistic analogy is meant to establish how each culture sets particular parameters that, once established, make our understanding of another culture's settings as incomprehensible as Chinese is to an American speaker or Greek is to someone speaking Afrikaans. Appreciating that universality exists is certainly not a dismissal of the importance of religion or government or education. It is an acknowledgment that some aspects of our moral psychology have nothing at all to do with these explicitly created institutions and, in fact, are immune to them. This should not cause dismay to those who believe in these institutions nor should it cause the moral grammarians to gloat. Rather, we all should be comforted by the fact that at the end of the day, we are united by a universal moral grammar that may help us see eye to eye, and see where our evolved intuitions can lead us astray in a world that is continuously changing.

The work discussed here has been conducted by Marc D. Hauser in collaboration with two graduate students, Fiery Cushman and Liane Young, and with helpful comments from other members of the Cognitive Evolution Lab at Harvard. The Internet survey can be found online at moral.wjh.harvard.edu.