# Resisting the Linguistic Analogy: A Commentary on Hauser, Young, and Cushman

(In W. Sinnott-Armstrong (Ed.) (2008). *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity.* MIT Press.)

## Jesse J. Prinz jesse@subcortex.com

In the 18<sup>th</sup> century, it was popular to suppose that each human capacity was underwritten by a specialized mental faculty. This view was championed by phrenologists well into the 19<sup>th</sup> century, and then rejected by behaviorists in the early 20<sup>th</sup> century. In contemporary cognitive science, faculties are back in vogue, due largely to the influence of Noam Chomsky's work on universal grammar. In addition to the language faculty, contemporary researchers also postulate dedicated faculties for reasoning about psychology, math, physical objects, biology, and other domains that look like a list of university departments. Conspicuously absent from this list is a faculty dedicated to morality. This was the most popular faculty of all, back in the days when men wore white wigs, and it is long overdue for a comeback. In their stimulating chapter Marc Hauser, Liane Young, and Fiery Cushman postulate an innate system dedicated to morality, and they speculate that it is interestingly similar to Chomsky's universal grammar. Related views have also been defended Mikhail (2000), and Dwyer (1999), and Rawls (1971). Hauser et al. do much to sharpen the language analogy, and they also bring recent empirical findings to testify in its defense. I applaud these contributions. Their hypothesis deserves serious attention, and their experimental findings provide data that any naturalistic theory of moral psychology must accommodate.

That said, I think it is premature to celebrate a victory for the moral faculty. There are alternative explanations of the current data. Instead of deriving from an innate moral sense, moral judgments may issue from general-purpose emotion systems and socially transmitted rules. Like art, religion, and architecture, morality might be an inevitable byproduct of other capacities rather than an ennobling module. In what follows, I raise some questions about the linguistic analogy, I express some doubts about the innateness of a moral faculty, and I sketch a non-nativist interpretation of the experimental findings that Hauser et al. present. I do not take my objections to be decisive. Hauser et al. may be right. Rather, I offer a non-nativist alternative with the hope that the dialogue between faculty theorists and their detractors will help guide research.

# 1. The Linguistic Analogy

Hauser et al. believe that there are a number of similarities between morality and language. They say that both capacities:

- have an innate universal basis,
- are vulnerable to selective deficits,
- exploit combinatorial representations,
- and operate using unconscious rules.

If all four points of comparison are true, then there is indeed an analogy to be drawn between language and morality. I am skeptical about each point, but before making that case, I must enter further point of concern. Notice capacities other than language, such as vision and motor control, are underwritten by mechanism that have each of the items on this list. Thus, the "language analogy" might equally be called the "vision analogy" or the "motor analogy." By drawing an analogy with *language* in particular, Hauser et al. are implying further points of comparison that may not hold up when all the evidence is in. Consider five potential disanalogies.

First, language has a critical period. This may be true of some perceptual systems too, but studies of, e.g., vision restoration late in life, suggest that language may be somewhat unusual in this respect. We don't know if there is a critical period for morality, but there are anecdotal reasons for doubt. Case studies of children who were raised in isolation, such as Genie or the wild boy of Aveyron, do not report profound moral deficits. Moreover, people can also acquire new moral values late in life, as happens with religious conversion, feminist consciousness raising, and a general trend from liberal to more conservative values that can be traced across the lifespan. Unlike language, learning a second morality does not seem fundamentally different than learning a first.

Second, language is usually learned in the absence of negative or corrective feedback. Is this true in the case of morality? Arguably not. Children are punished for making moral mistakes: they are reprimanded, socially ostracized, or even physically disciplined. Children also hear adults expressing negative moral attitudes towards numerous events. Of course, kids are never explicitly taught that it's worse to push people off of footbridges then to killing by switching the course of a speeding steam engine, but these specific rules may be extrapolated from cases acquired through explicit instruction, as I will suggest below.

Third, according to leading theories of grammar (e.g., Chomsky's Government and Binding theory), linguistic rules are parameterized: they have a small set of possible settings that are triggered by experience. Hauser et al. explicitly endorse this view for morality, but it's not clear what the parameters are supposed to be. Consider opposing moral systems, such as liberalism and conservatism. It doesn't look like the conflicting values are simply different settings on the same basic formation rules. Where linguistic parameter settings correspond to structural variations in how to combine primitives, variation in moral values does not seem to be structural in this sense. Consider the moralized political debate social welfare: should governments give aid to those in need, or should the distribution of wealth be determined entirely by what individuals manage to attain in the free market? This question concerns a conflict between principles of equality and equity, rather than a conflict between alternative settings for the same basic principle. Or consider the debate about capital punishment; the two positions are dichotomous (pro or con), and they stem from different conceptions of punishment Similar considerations apply to debates about gender (retribution and deterrence). equality, gun control, and the moral permissibility of imperialism. These differences cannot be treated as parametric variations, except by trivializing that idea—i.e, treating each contested policy as a parameter in its own right, which can be switched on or off. Haidt and Joseph (2005) argue that political conservatives have moral systems that contain categories of rules (e.g., rules about hierarchy, honor, and purity) that are not part of liberal morality, rather than mere variations on rules of the kind liberals share. Of course, there are some classes of rules that crop up in most moral systems, such as prohibitions against harm, but the variations in these rules are open-ended rather than parametric. Who may you harm? Depending on the culture, it can be an animal, a child, a criminal, a woman, a member of the out group, a teenager going through a right of passage, a person who is aggressing against you, an elderly person, and so on. The range of exceptions is as varied as the range of possible social groups, and there is equal variation in the degree to which harm is tolerated (brief pain, enduring pain, mutilation, disfigurement, death). If moral rules were parameterized, there should be less variation.

Fourth, when two languages differ in grammar, there is no tendency to think one grammar is right and the other one wrong. We never start wars to snuff out people who place nouns before adjectives. In contrast, participants in moral conflicts assume that their values are the only acceptable values.

Fifth, language uses various levels of representation: phonology, syntax, and semantics, each of which may subdivide into further levels. There doesn't seem to be an analogous range of moral levels of representation.

Of course, Hauser et al. can concede these points of contrast and restrict their analogy to the four similarities laid out above. That would weaken the language analogy, but it wouldn't undermine it. But each of the four alleged similarities is itself subject to doubt. Let's have a look.

Do moral rules operate unconsciously? To support this claim, Hauser et al. show that people are bad at justifying their moral judgments. But this is evidence for unconscious rules only if we think those rules should take the form of justifying principles. Suppose that moral rules take the form of simple injunctions: It's horrible to intentionally kill someone; it's pretty bad to let someone die; we have special obligations to people close to us; incest is seriously wrong; steeling is wrong too, but not as bad as physically harming; and so on. These rules are certainly accessible to consciousness. They are usually much more accessible than the rules of language.

Are moral rules combinatorial? This is a bit more complicated. As Hauser et al. point out, we certainly need a combinatorial system for categorizing actions. But notice that action categorization is something we do quite independently of morality. Our capacity to tell whether something was done intentionally, for example, operates in nonnonmoral contexts, and individuals who lack moral sensitivity (such as psychopaths) are not impaired in recognizing actions or attributing intentions. Psychopaths can recognize that someone is intentionally causing pain to another person. Moral rules take these combinatorial, non-moral representations of actions as inputs and then assign moral significance to them. The distinctively moral contribution to a rule such as killing is wrong, is not the representation of the action (killing), but the attitude of wrongness. It's an interesting question whether moral concepts, such as *wrong* have a combinatorial structure; they may. But, by focusing on the combinatorial structure of action representations, Hauser et al. fail to show that representations specific to the moral domain are combinatorial.

Is morality vulnerable to selective deficits? I just mentioned psychopaths, who seem to have difficulty understanding moral rules. This can be inferred from the fact that psychopaths don't exhibit moral emotions, they engage in anti-social behavior, and they fail to distinguish between moral and conventional rules (Blair, 1995). But psychopathy is not a selective deficit in morality. Psychopaths have other problems as well. They seem to suffer from a general flattening of affect, which also affects their ability to recognize emotional facial expressions and to recognize emotion intonation in speech (Blair et a., 2002). Psychopaths may also suffer from a range of executive disorders. They tend to be disinhibited, and they make cognitive errors as a result (e.g., errors on maze tasks) (Sutker et al., 1992). The moral deficit in psychopathy may result from their general emotion deficit. With diminished negative emotions, they don't experience empathy or remorse, and that leads them to be dangerously indifferent to the well-being others. If this analysis is right, then psychopathy is a domain-general problem with moral repercussions. I know of no case in the clinical literature in which morality is impaired without co-morbid impairments of other kinds, most notably emotional impairments.

Is morality innate and universal? This question requires a bit more discussion.

#### 2. Moral Judgments and Innateness

Elsewhere I have defended the claim that morality is not innate (Prinz, this volume; forthcoming a; forthcoming b). I will not rehearse all my arguments against nativism here, but I want to highlight some issues of contention that can help focus the debate.

To decide whether moral judgments are innate, we need a theory of what moral judgments are. Hauser et al. review several different accounts of moral judgment, or, at least, or how moral judgments relate to reasoning and emotion in information processing. On one model, which I'll call Reasons First, things proceed as follows: we perceive an event, then reason about it, then form a moral judgment, and that causes an emotion. On an Emotions First model, the sequence goes the other way around: we perceive an event, then form an emotion, that causes a moral judgment, and then we reason about it. On their view, neither of these is right. Instead, they favor an Analysis First model: we first perceive an event, and then analyze in terms of component features such as INTENTION, AGENT, RECIPIENT, HARM; this leads to a moral judgment, which can then give rise to emotions and reasoning. I think Hauser et al. are absolutely right that moral judgment typically requires action analysis, but they are wrong to deny that other theories leave this part out. One cannot make a moral judgment about an event without first categorizing that event. Only a straw version of the Reasons First and Emotions First models would leave out some kind of action analysis. Still, there are two important differences between the Hauser et al. model and these others. First, for Hauser et al., action analysis in not done by a domain general mechanisms that is used for categorizing actions; rather, it is done by the moral faculty, which analyzes actions using features that may be proprietary to making moral assessments. Second, for Hauser et al., both emotion and reasoning occur after moral judgments are made. So their model is a genuine alternative to these others.

Of these three models, I am most sympathetic to Emotions First, but my view pushes that approach even farther. On the Emotion First model that Hauser et al. consider, emotions *cause* moral judgments. Jonathan Haidt (2001) favors such a view, but he never tells us exactly what moral judgments are. For example, he doesn't tell us what concept is expressed by the word "wrong"? Hauser et al. don't tell us the answer to that question either. I think the concept expressed by "wrong" is constituted by a sentiment. A sentiment is the categorical basis of a disposition to experience different emotions. The sentiment that constitutes the concept wrong disposes its possessor to feel emotions of disapprobation. If I judge that stealing is wrong, that judgment is constituted by the fact that I have a negative sentiment towards stealing—a sentiment that disposes me to feel angry at those who steal and guilty if I myself steal. On any given occasion in which I judge that something is wrong. I will likely experience one of these emotions, depending on whether I am the author of the misdeed or someone else is. (And likewise for other moral concepts.) Thus, in place of the Emotions First model, on which emotions cause moral judgments, I favor an Emotion Constitution model, according to which emotions constitute moral judgments. More fully elaborated, I think moral judgment involves the following sequence: first, we perceive an event and categorize it; if that event-type matches one towards which we have a stored sentimental attitude, the event triggers the relevant emotion in me (e.g., guilt if it's my action and anger if it's yours). The resulting mental state is a representation of perception of an action together with a sentimental toward that action, and this complex (action representation plus emotion) constitutes the judgment that the action is wrong. The moral judgment is not a further stage in processing following on the heels of the emotion, but is constituted by the emotion together with the action representation. After that, I might reason, or put my judgment into words, or re-assess the case and adjust my sentiments, and so on.

I can't defend this theory of moral judgment here. The evidence is both philosophical and empirical. The empirical evidence is the same as the evidence used to support the Emotions First model: emotions seem to occur when people make moral judgments; emotion induction alters moral judgments; and emotion deprivation (as in the case of psychopathy) leads to deficits in moral judgment. But the Emotion Constitution model has an advantage over the Emotion First model: it is more parsimonious. Rather than saying moral concepts are mental entities that are caused by moral emotions, I say they are constituted by moral emotions. This fits with the pretheoretical intuitions. A person who feels guilty or outraged about some event can be said, *in virtue of those emotions emotions*, to have a moral attitude about that event. This suggests that emotions constitute moral attitudes. Hauser et al. will presumably disagree. For present purposes, I simply want to explore what implications this approach to moral judgment has for nativism.

If moral judgments are constituted by emotions, then the question of whether morality is innate boils down to the question: how do we come to have the emotions we have about things such as stealing, killing, cheating, and so on? A nativist will propose that we are innately disposed to have these emotions in virtue of domain-specific principles (which may be parameterized). Here's a non-nativist alternative. Suppose that a child who has no moral attitudes or moral faculty engages in a form of behavior that her caregivers dislike. The caregivers may get angry at her, and they may punish her in some way. For example they might scold her or withdraw love and affection. Children rely on the affection of caregivers, and when punished those all-important attachments are threatened. The emotion elicited by threats to attachment is sadness. Thus, a child who misbehaves will be led to feel badly. Over time, she will associate that feeling of sadness with the action itself; she will anticipate sadness when she considers acting that way again. Once the child associates sadness with that action, we can say she feels regret, remorse, or even guilt about it. These moral emotions can be defined as species of sadness directed at courses of action. The main difference between ordinary sadness and guilt is that guilt promotes reparative behavior. Such behaviors need not be innate. They are a natural coping strategy for dealing with cases where you have angered another person. The child who is punished will also come to have the same anger dispositions as those that punish her. Children are imitative learners. If a child sees her parents get angry about something that she does, she will feel sad about it, but she will also come to feel angry at other people when they engage in that behavior. She will copy her caregiver's reactions. This will also allow children to acquire moral rules concerning behaviors that they have never attempted; such as prohibitions against murder and rape. When such behaviors are mentioned by caregiviers, there is almost always an expression of emotion. When we mention things that we morally oppose, we do not conceal our emotions. Children imitatively pick up these attitudes. Notice that this story explains the transmission of moral rules by appeal to domain-general resources: children must be able to categorize actions, they must experience sadness when punished, and they musty be disposed to imitate anger and other negative emotions expressed by caregivers. If a child has these capacities, she will learn to moralize. She does not need an innate moral sense.

This developmental just-so story is intended as a possible explanation of how one could learn moral rules without having an innate moral faculty. If moral judgments are sentimental, then moral rules are learnable. But it is one thing to say that moral rules are learnable and another thing to say they are learned. After all, we could be born with innate moral sentiments or sentimental dispositions. Just as we are biologically prepared to fear spiders, we might be biologically prepared to feel angry and guilty about various courses of action. We need a way of deciding whether moral rules are innate or acquired. One way to approach this question is development. Do children acquire certain moral rules more easily? Are others impossible to acquire? Are certain moral rules learned without punishment, or other kinds of social interaction that condition emotional responses? I think these are all important open questions for research. I do think that there is extensive evidence for the claim that punishment plays a central role in moral education (Hoffman, 1983), and that leads me to think that moral nativism will be difficult to defend by appeal to a poverty of the stimulus argument, as I mentioned above. I also think that the wide range of moral rules found cross culturally suggests that children can acquire moral attitudes towards just about anything. But both of these observations are anecdotal, and it is crucial at this stage to systematically search for innately prepared moral rules.

# 3. Trolley Cases

In suggesting that morality may not be innate, I don't want to deny that are innately disposed to engage in some forms of behavior that are moral praiseworthy. Perhaps helping behavior, reciprocal altruism, and various forms of peacemaking are species typical in the hominid line. But there is a difference between behaving morally and making moral judgments. My hypothesis is that people are not innately equipped with a faculty of moral judgment. Moral concepts, such as right and wrong, are acquired from domain general mechanisms. The fact that we are innately disposed to do some praiseworthy things is no more evidence for innateness of a moral sense than is the fact that we are disposed to take care of our young. Laudable behavior can exist without the

capacity to praise it as such. One of the exciting features of Hauser et al.'s research program is that they are directly investigating moral judgments, rather than morally praiseworthy behavior. Their research on trolley cases can be interpreted as an argument for innate moral judgments.

Here's how I interpret that argument. There are moral judgments about moral dilemmas that are very widespread, homogenous across different demographics, and demonstrable across cultures. These judgments do not seem to be learned through explicit instruction and they do not seem to be based on consciously accessible reasoning processes. Together, this pattern is consistent with the conclusion that the judgments issue from an innate moral faculty. It's not a demonstrative argument, of course, but it's a reasonable argument to the best explanation—or at least it would be, if there weren't other equally good non-nativist explanations available.

Here's how a non-nativist might account for the data. On my view, there is a moral rule of the form: intentionally taking another person's life is wrong. This rule consists of a domain general action representation (intentionally taking a person's life) and a sentiment (which disposes one to feel angry or guilty if a person is killed by someone else or by oneself). The non-nativist needs to explain how such a rule could come about without being hard-wired. That does not look like an insuperable challenge. Societies that allow killing, at least within the in-group, are not very stable. In very small-scale societies, built around extended kin groups, there may not be a need for any explicit rule against killing. We rarely have motives to kill our near and dear, especially if we feel a sense of attachment to them. But, as societies expand to include many nonnonrelatives, relatives, pressure arises to introduce killing norms, and that probably doesn't take much work. If you try and kill your neighbor, he and his loved ones will get pretty miffed. Other members of the community, afraid for their own security, may get upset too, and they will try punish you or banish you. Thus, aggression against others, naturally elicits strong reactions, and those reactions condition the emotions of the aggressor. Knowing that aggression can lead to alienation and reprisal, you resist. When you think about aggressing, you feel anticipatory guilt, and, when you imagine others aggressing, you get angry about the harm they will do. Thus, we don't need innate strictures against killing, because the natural non-moral emotions that are elicited by acts of aggression will instill the sentiments that constitute moral disapprobation. The rules against killing may, at first, be limited to the in-group, because aggression against more distant strangers may go unpunished by the local community. But, when communities become more transient, more diverse, or more dependent on others for trade, strictures against killing generalize, because harming distant strangers can be perceived as a potential threat to members of the local group.

So much for the genealogy of norms against killing. The non- nativist also needs to explain helping norms. Most of us think we should help people in need if we can do so at little personal cost. Is this an innate rule? Not necessarily. It could easily emerge through cultural evolution, because helping confers obvious advantages. If I join a group whose members will help me when I am in need, I will fare better, than if I join a group of selfish people. But helping always introduces free rider problems. How can I be sure that people in my community will help me? Game theoretic models suggests that the best solution for coping with free riders is punishment. If I penalize people for being unhelpful, then they will be more likely to help in the future. Punishment leads people to feel guilty about free riding and angry at other free riders. Thus, when unhelpful individuals are punished, emotions are conditioned, and a moral attitude in born. In sum, I think the social and emotional consequences essentially guarantee that most societies will end up with moral rules about killing and helping. Non-violent cooperation may be a precondition to stability in large populations. But these rules about killing and rules about helping may differ from each other in one respect. Several factors are likely to make killing norms stronger than helping norms. First, in cultural evolution, prohibitions against killing are more vital than prohibitions against unhelpful behavior, because a group whose members kill each other will fare worse than a group of members who go out of their way to help each other. Second, helping also carries more personal cost than refraining from killing. Third, acts of aggression naturally elicit fear and anger, so it is easier to inculcate strong sentiments towards killing. Collectively, these factors essentially guarantee that sentiments towards killing will be stronger then sentiments pertaining to helpful and unhelpful behavior. If the Emotion Constitution model of moral judgment is right, this difference in sentimental intensity is tantamount to a different in the strength of the respective moral rules.

I have been arguing that we can account for norms about helping and killing without suppose that they are innate. Once they are in place, they can guide behavior, and, on occasion, they will come into conflict. When this happens, there are two factors that will determine which rule will win. One factor is the extent to which actions in the event under consideration can be construed as instances of killing, on the one hand, or helping. Failure to conform to paradigm cases of either will diminish the likelihood that we will apply our rules about killing and helping. If some course of action is only a borderline case of killing, we may apply our killing rule with less force or confidence. For example, suppose someone causes a death as a side effect of some other action. This is not a paradigm case of killing. In terms of cultural evolution, groups have greater interest in condemning people who form direct intentions to kill, rather than people who kill as a side effect, because the person who will kill intentionally pose a greater threat. Killing without the explicit intention to kill is a borderline case of the rule. The other factor is emotional intensity. For example, if we can help a huge number of people, our helping rule may become emotionally intense. In some cases, emotions may be effected by salience: if attention is drawn to an act of helping or killing, the corresponding rule will be primed more actively, and the emotions will be felt more strongly. Now at last, we can turn to the trolley cases presented by Hauser et al. These cases are interesting because they pit helping norms against killing norms. We can now see whether the nonnativist, emotion-based theory can explain the results. In the first case case, Frank is on top of a footbridge, and can push a man into the path of a trolley, thereby saving five people further down on the track. Only 11% of subjects think its okay to push the man. One explanation is that this is a paradigm case of killing, and the killing rule is, all else being equal, more emotionally intense than the helping rule. It's also a very salient case of killing, because subjects have to imagine Frank pushing someone, and the thought of physical violence attracts attention and increases emotion. In a second case, Denise can pull a lever that will guide a trolley down an alternate track, killing one person, rather than allowing it to kill the five people on the track it is currently on. Here 89% say it's permissible to pull the lever. The numbers change because this is not a paradigm or emotionally intense case of killing. The person who is killed is not physically assaulted, and Denise does not need to form the intention: I want to cause that guy's death.

The next case is a bit puzzling at first. Like Denise, Ned can pull a lever that will send a train on a different track, killing one, rather than five. But, unlike the Denise case, the track in Ned's case is a loop, that would reconnect with the original track and kill the five people were it not for the fact that the guy on the alternate track is heavy enough to stop the trolley in its tracks. In this situation, only 55% of subjects think Ned is permitted to pull the lever, killing one and saving five. Why would the minor addition of a looping track change permissibility judgments from the Denise case? The answer may be salience. When we imagine a person being used to stop a trolley in its path, the imagery is more violent and more emotionally intense. It is also a more paradigmatic case of killing, because Ned has to explicitly form the intention that the person be crushed, otherwise the train wouldn't stop.

Hauser et al.'s final case in a slight variant on the Ned case. Here, Oscar can pull a lever that will send a train on a loop track that is obstructed by a large weight; the weight will prevent the train from rejoining the original track where it would kill five, but, unfortunately, there is a man standing in front of the weight who will be killed if the lever is pulled. 72% of subjects think this is permissible. These permissibility ratings are higher than in the Ned case, because it is a less paradigmatic case of killing: the death in the Oscar case is an accidental byproduct of sending the train into the weight. There is just one remaining question: why are the permissibility ratings in the Oscar case slightly lower than in the Denise case? The answer may involve salience. In the vignettes, the solitary man in the Oscar case is introduced with a 20-word sentence, and the solitary man in the Denise case is introduced with 10 words. In the Oscar case that man is crushed between the train and the weight, and in the Denise case, he is killed the same way that the five people on the other track would have been killed. Thus, the Oscar case draws extra attention to the victim. These explanations are sketchy and tentative. I offer them to illustrate a simple point. If one can tell a non-nativist and sentimentalist story about moral rules pertaining to killing and helping, there are resources to explain intuitions about trollev cases. Without ruling out this alternative account, Hauser et al.'s argument for nativism loses its force. At this stage, it's fair to say that both the nativist and the non-nativist accounts are in embryonic stages of development, and both should be considered live options as we investigate the origin of our capacity to make moral judgments.

The account that I have been proposing leads to some predictions. The first is consistent with Hauser et al.'s account, the second is slightly harder for them to accommodate, and the third is is more naturally predicted by my account. First, I think that moral rules contain representations of actions, and these representations may take the form of prototypes or exemplars (a typical murder, for example). I predict that the moral judgments will weaken as we move away from these prototypes. Hauser et al. may agree.

Second, I think that helping and harm norms are socially constructed to achieve stability within large groups, and consequently, there may be subtle cultural differences as a function of cultural variables. For example, consequentialist thinking may increase for groups that are highly collectivist (hence more focused on what's best for the collective), for groups that are engaged in frequent warefare (hence more desensitized to killing), and for groups that are extremely peaceful (where norms against killing have

never needed to be heavily enforced). In highly individualist societies, there is less overt focus on helping behavior, and consequentialist thinking may diminish. Likewise in highly pluralistic societies; pluralism promotes the construction of strong rules against killing, because such rules are often needed to ensure peace in diverse groups. Hauser et al. report on some cross-cultural work, but there are two limitations of the data they report. First, as they note, their non-American subjects understand English and have access to computers, so they are probably similar to us. Second, Hauser et al. do not report the actual percentages for their cross-cultural samples; so even if every tested culture tended to say Frank's behavior is less permissible than Denise's, the actual percentages who hold that dominant view may differ. It is important to note that Hauser et al. can allow variation in moral judgments. The language analogy predicts that principles will have parameters that get set differently in different contexts. My worry is that this is the wrong kind of variation. In language, switching parameters results in differences that are qualitative and arbitrary. The differences that I am imagining are quantitative and tailored to fit cultural variables. That is suggestive of learning, rather than innateness.

Third, the Emotion Constitution Model predicts that manipulation of emotions should influence judgments on trolley dilemmas. By making one of the two courses of action more salient, more violent, more personal, or more emotionally evocative in some other way, one should be able to alter the permissibility ratings. Psychopaths should not be influenced to the same degree by emotional manipulations. Such findings would count against Hauser et al.'s non-affective theory of moral judgment, and they would also count against the view that moral judgments are driven by domain-specific (or at least encapsulated) mechanisms.

If these predictions pan out, they add support the Emotion Constitution model. That model is compatible with nativism, but it also lends itself to a plausible non-nativist account of how we come to acquire moral rules. In this commentary, I haven't provided strong evidence for the non-nativist view or against the view favored by Hauser et al. Rather, my goal has been to suggest that, at this early stage of inquiry, several models remain compatible with the evidence. Hauser et al. would undoubtedly agree, and, in the coming years, we will need to find ways to test between these options. Let me sum up with a few questions for Hauser et al. that highlight places where their model and my alternative come apart. Why think that the analyses of action that precede moral judgment are carried out by a domain-specific moral faculty? Why think that emotions arise as consequences of moral judgments rather than causes or constituent parts? Why think that moral principles are innate rather than learned solutions to problems facing all cultures? And what is it about language, as opposed to any other faculty, that sheds light on our moral capacities? Hauser et al. have embarked on an important research program, and the linguistic analogy has been a valuable source of inspiration. My hunch is that it will eventually prove more productive to drop that analogy and adopt a model that places greater emphasis on learning. For now, we can make most progress by keeping both approaches on the table.

# References

- Blair, R.J.R. (1995). A Cognitive Developmental Approach to Morality: Investigating the Psychopath. *Cognition*. 57, 1-29.
- Blair, R.J.R., and Mitchell, D.G.V., Richell, R.A., Kelly, S., Leonard, A., Newman, C. and Scott, S.K. (2002). Turning a Deaf Ear to Fear: Impaired Recognition of Vocal Affect in Psychopathic Individuals. *Journal of Abnormal Psychology*, 111, 682–686.
- Dwyer, S. (1999). Moral competence. In K. Murasugi and R. Stainton (Eds.), *Philosophy and Linguistics Linguistics*. Westview Press.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108 108*, 814-834.
- Haidt, J. and Joseph, C. (2005). Intuitive ethics. Daedalus Daedalus, \*\*, , \*\*.
- Hoffman M.L. (1983). Affective and cognitive processes in moral internalization. In E.T.
- Higgins, D.N. Ruble, W.W. Hartup (eds.), *Social Cognition and Social Development: A Sociocultural Perspective Perspective*, pp. 236-74. Cambridge: Cambridge University Press.
- Mikhail, J. (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.
- Prinz, J. J. (forthcoming a). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Prinz, J. J. (forthcoming b). Against moral nativism. In D. Murphy (ed.) \*\* Stich Volume\*\*\*.
- Prinz, J. J. (this volume). Is morality innate?
- Rawls, J. (1971). A Theory of Justice. Cambridge, MA: Harvard University Press.
- Sutker, P. B., Moan, C. E., & Swanson, W. C. (1972). Porteus Maze test qualitative performance in pure sociopaths, prison normals, and antisocial psychotics. *Journal of Clinical Psychology*, 28 28, 349–353.