

On Misreading the Linguistic Analogy:**Response to Jesse Prinz and Ron Mallon**

Marc D. Hauser, Liane Young, and Fiery Cushman

Oscar Wilde noted “Always forgive your enemies - nothing annoys them so much.” Before we forgive our critics, however, we thank Prinz and Mallon for their thoughtful comments, and for taking the linguistic analogy as a serious proposal amidst the current excitement at the interface between moral philosophy and moral psychology. What we forgive is their targeted comments on several issues that are either irrelevant to the linguistic analogy or premature given that we know so little about the nature of our moral psychology. Some of the confusion is undoubtedly due to our own exposition, and some to the rapid pace of theoretical and empirical developments that have emerged since we submitted the final draft and received the commentary.

We begin by clarifying the main goals of the linguistic analogy, including, most importantly, its unique set of empirically tractable questions and challenges. Our hope is that this response, guided by Prinz and Mallon’s comments, serves as the next installment on a much larger project that, we can all agree, will yield interesting results irrespective of the strength of the analogy. The reason for this is simple: until the questions that emerge from the analogy are taken seriously, and pitted

against the alternatives, we will have only a weak understanding of the mature state of moral knowledge, how it is acquired within the individual and species, and the extent to which it relies upon domain-specific machinery. In this sense, we see the arguments generated in our target essay, and developed more fully elsewhere (Dwyer, 2004; Hauser, in press; Mikhail, 2000), as analogous to the minimalist program in linguistics (Chomsky, 1995, 2000): a set of fascinating questions with ample room for movement on both theoretical, empirical, and methodological fronts.

For a novel research program to breathe, it is important that its claims be properly understood and that challenges be targeted at the proper level. Let us start then by highlighting two important points of agreement: both Prinz and Mallon 1) endorse our research program focused on the cognitive systems responsible for generating the basic representations that serve as input to the process of moral judgment, and 2) support our position that these systems operate over the representations of actions, intentions, causes and consequences. By supporting these two points they at least implicitly support a third which, we submit, follows: some moral principles are formulated over the core representations that enter into our moral judgments. The primary thrust of the linguistic analogy is to study these systems and bring them to the attention of philosophers and psychologists. It is in this spirit that we turn next to a more detailed look at the linguistic analogy, pinpointing what we

perceive as its central assumptions and predictions, together with a body of relevant data. Along the way we point out some of the challenges raised by Prinz and Mallon, including the non-nativist alternative based on emotions, and emphasize the need to posit an innate, dedicated moral organ.

Both Prinz and Mallon attribute to us the view that the cognitive systems responsible for generating basic representations used in moral judgment are in fact specific to the domain of morality. This is not our view—indeed, it should have been clear that we hold the opposite position. Moral judgment depends on a wide range of representational inputs generated by cognitive systems adapted for and typically engaged in entirely different functions. Analogous cognitive mechanisms support linguistic competence without being specific to the domain of language. To clarify, take the rather simple phenomenon of speech perception. Although the last fifty years of research has largely assumed that we are endowed with a dedicated neural system for processing speech, neuroimaging studies with normal subjects, together with comparative and developmental studies of other animals and infants, suggests that much of speech perception may derive from very general, and ancient auditory mechanisms. For example, a recent study by Vouloumanos, Hauser & Werker (unpublished manuscript) showed that neonates less than 48 hours evidenced no preference for human speech over rhesus

monkey vocalizations. Similarly, comparative studies of human adults, infants, and cotton-top tamarin monkeys revealed no difference in the capacity to use transitional probabilities to segment a continuous stream of speech (Hauser, Newport, & Aslin, 2001). These results suggest that early stages of speech perception and segmentation are not mediated by processes that are specific to the domain of language.

Though we explicitly recognize the role of domain-general mechanisms, we are nonetheless committed to the existence of some cognitive mechanisms that are specific to the domain of morality. These we term the moral faculty. These systems are not responsible for generating representations of actions, intentions, causes and outcomes; rather, they are responsible for combining these representations in a productive fashion, ultimately generating a moral judgment. Our thesis is that the moral faculty applies general principles to specific examples, implementing an appropriate set of representations. We refer to these principles as an individual's 'knowledge of morality', and by analogy to language, posit that these principles are both unconsciously operative and inaccessible.

Mallon notes that we must distinguish between a theory that can adequately account for the pattern of people's moral judgments and a theory that is actually instantiated in people's heads. We fully agree, especially since this captures the parallel distinction in linguistics. To be

precise, we must distinguish between a set of principles that are descriptively consistent with people's moral judgments and the principles that people in fact carry around in their heads, doing the work of adjudicating between moral rights and wrongs. As Mallon correctly intuited, we are aiming at principles in the head. But the first step, of course, is to determine the set of principles at the descriptive level.

Consider the following example as an illustration of how first to identify the set of descriptive principles that are operative in guiding moral judgment and then to investigate the extent to which these principles are expressed in the course of justification. In a recent paper (Cushman, Young, & Hauser, 2006) focused on the relationship between operative and expressed principles, we develop the argument that a three-pronged approach is necessary to assess whether particular principles mediate our moral judgments and whether these principles serve as the basis for our justifications. *Prong 1*: Develop a battery of paired dilemmas that isolate psychologically meaningful and morally relevant, principled distinctions. *Prong 2*: Determine whether these targeted principles guide subjects' moral judgments. *Prong 3*: Determine whether subjects invoke these principles when justifying their moral judgments. With this approach, we explored three principles:

Action Principle: Harm caused by action is morally worse than equivalent harm caused by omission.

Intention Principle: Harm intended as the means to a goal is morally worse than equivalent harm foreseen as the side-effect of a goal.

Contact Principle: Harm involving physical contact with the victim is morally worse than equivalent harm involving no physical contact.

Based on a sample of approximately 300 subjects, largely from English-speaking, Western countries, analyses revealed support for the three targeted principles in 17 out of 18 paired dilemmas. That is subjects judged harm caused by action as worse than omission, intended harm as worse than foreseen harm, and harm involving contact as worse than with no contact. When we turned to justifications, 80% of subjects recovered the key distinction for the action-omission cases, 60% for the contact-no contact cases, and only 30% for the intended-foreseen cases. This pattern suggests that the intended-foreseen distinction is operative, but results in an intuitive judgment. The other principles are also operative but appear to be at least accessible to conscious awareness.

Are the descriptive principles targeted in this study isomorphic to the domain-specific principles that constitute an individual's moral

knowledge? At present we cannot say. We know that these principles are descriptively adequate to capture the observed pattern of subjects' moral judgments, but it remains a viable possibility that they exert their influence during the generation of the relevant representations that are external to and feed into moral judgment. Of course, a direct implication of the view that these principles are not specific to morality is that they influence judgments and behaviors outside the moral domain. Identifying non-moral analogues of these descriptive principles—if indeed they exist—is an important area for future research.

Thinking about the moral faculty from this perspective leads us directly into Mallon's point that evolution may have created particular biases that set initial conditions on the valenced responses. Consider sex, and the extent to which degrees of genetic relatedness matter. An agent INTENDS/DESIRES to +/-SEXUAL INTERCOURSE with X_r , where X is some sexual partner and r is their degree of genetic relatedness to the agent. If we ask whether sexual intercourse is morally permissible with X , the answer depends on r . Evolution appears to have set up a bias, in the sense that r values between .125 and .5 are generally coded as -SEXUAL INTERCOURSE. This may be the default setting or bias, open to modification (to some extent) by the local culture. Again, the initial valence settings may have been established on the basis of their statistical effects (e.g., the probability that mating with parents and siblings will

reduce fitness), and only later, hooked into the emotions as reinforcing agents. In sum, we completely agree with Mallon that evolution has set us up with strong biases. These biases may enter into moral judgments, and at this point, we are agnostic on whether they figure into moral competence or performance.

To summarize thus far, we propose, and Prinz and Mallon agree, that a deeper understanding of the sources of our moral judgments requires further research into the nature of our representations of actions, intentions, causes and consequences. The system involved in generating such representations is not specific to the moral domain. In parallel to language, however, individuals possess knowledge of morality that is comprised of domain-specific moral principles operating over these representations. Though we are only at the earliest stages of this research program, our empirical studies suggest a methodology to determine candidate principles for domain-specific moral knowledge. Whether the descriptive principles that capture patterns of moral judgment in fact characterize features of the moral faculty or features of the cognitive systems that feed into the moral faculty is presently unknown, but, we submit, not unknowable.

What we wish to stress is that the linguistic analogy provides a substantive foundation for constructing testable hypotheses and collecting the relevant data. For example, as a theory, it demands a proper

descriptive account of the mature state of moral knowledge. Until we understand our moral psychology at this descriptive level, including some subset of its principles, it is virtually impossible to make progress on other fronts, including especially, issues of moral acquisition (explanatory adequacy in Chomsky's terms), domain-specificity, characteristic neural breakdown, and evolutionary origins. That is, we need to understand the nature of our mature subject's moral knowledge before we can ask how it evolved, develops, and is instantiated in neural tissue.

A thorough characterization of moral knowledge is particularly critical to adjudicate between nativist and empiricist claims. For example, Prinz states that he doubts there is a critical period for morality in the same way that there is for language or that learning a second moral system is like learning a second language. But we are only able to determine that there is a critical period for language because we have a relatively deep understanding of the principles underlying the mature state of linguistic knowledge and thus, can see what happens to the externalization of such knowledge in expressed language as a function of severe developmental isolation. Furthermore, we are only able to contrast native and second language acquisition because we understand *what* is being acquired. On the basis of a clearly characterized linguistic target, often articulated in terms of principles and parameters, we can state that native language acquisition is fast, effortless, untutored, and relatively

immune to negative evidence or correction. Second language acquisition is slow, effortful, tutored, vulnerable to negative evidence and correction. Surprisingly, no one has ever systematically compared the acquisition of native and second moral systems.

We end here with a discussion of the role of emotions in guiding our moral psychology and behavior. Though many of the questions that emerge from adopting the linguistic analogy have little or nothing to do with the emotions, our perspective puts into play a different way of looking at the role of emotions. To clarify, consider three ways in which emotions might enter into our moral judgments. First, an individual's emotional response to a particular circumstance might influence the representations he forms of the actions, intentions, causes and consequences associated with that circumstance. Second, an individual's emotional response to a particular circumstance might, itself, be among the representational inputs to the moral faculty. This characterization implies the existence of a domain-specific moral principle such as "if it produces negative affect, it is morally wrong." Finally, it is possible that emotion has no influence upon moral judgment but is only a product of it.

Prinz proposes "the Emotion Constitution model, according to which emotions constitute moral judgments". This corresponds most closely to our second possibility, but with some potential differences. On the one hand is the rather trivial and uncontroversial claim that moral

judgments are not synonymous with negative emotion. There are many instances in which we experience a negative emotion in the absence of moral disapproval (e.g. anger from stubbing a toe, disgust from seeing blood). On the other hand, Prinz appears to define moral judgment as a variety of negative emotion, such that the meaning of wrong is the feeling of wrongness. Stranding the problem here simply raises another: how does one determine wrongness in the first place? Prinz's solution is that "the concept expressed by 'wrong' is constituted by a sentiment ... (which is) the categorical basis of a disposition to experience different emotions." In essence, Prinz is describing a mechanism that has at its disposal some categorical basis (principles) that presumably operates over some set of representations and that outputs emotions that we label as 'right' or 'wrong' (moral judgments). Ironically, then, what Prinz calls a sentiment is apparently identical to what we call the moral faculty.

What the discussion above boils down to is that for both our perspective and the one Prinz favors, we are left with a binary choice: either emotion plays a role in moral judgments by shaping the representational input into the judgment mechanism (Prinz's sentiment, our moral faculty), or it is merely a consequence of that mechanism. This is an open and empirically tractable question that we have begun to explore. Let us illustrate with some recent patient data, acquired since our original submission.

In collaboration with Michael Koenigs, Daniel Tranel, Ralph Adolphs, and Antonio Damasio (in prep) we have explored the nature of moral judgments in six individuals with adult-onset bilateral damage to ventromedial prefrontal cortex (VMPC), an area noted for its critical role in linking emotion to decision making (Bechara, Damasio, Tranel, & Damasio, 1997). VMPC damage is associated with diminished autonomic and subjective response to passive viewing of emotionally charged pictures (Blair & Cipolotti, 2000; Damasio, Tranel, D., & DAMasio, 1990), recall of emotional memories (Tranel, Bechara, Damasio, & Damasio, 1998), contemplation of risky choices (Bechara et al., 1996), and consideration of counterfactual outcomes (e.g., regret) (Camille, Coricelli, Sallet, Pradat-Diehl, Duhamel, & Sirigu, 2004). We found that VMPC subjects were more likely to endorse personal or emotionally salient moral violations presented in hypothetical scenarios (developed by Greene et al. (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) than were comparison groups, including normals and brain damaged controls. More specifically, VMPC subjects were more likely to endorse violations that maximized aggregate welfare (e.g. throw a man off a bridge to save five others), resulting in heavily consequentialist judgments. There was no difference between VMPC subjects and comparison groups on either non-moral or impersonal moral scenarios, showing that many aspects of their decision

making systems are intact. A supplementary analysis of the personal moral scenarios showed that the difference between VMPC participants and comparison groups was restricted to the “difficult” as opposed to “easy” scenarios, as measured by uniformity of judgment within the comparison groups, showing further that even some judgments of emotional moral actions are intact. These analyses suggest that the effect of VMPC damage on moral judgment is both specific to its role in emotion processing and specific to scenarios for which there are no explicit adjudicating norms, that is, scenarios posing “difficult” moral dilemmas. In short, it appears that there may be an important role for emotion in shaping the representational inputs into the moral faculty.

These data bear on Prinz and Mallon’s concern about the notion of a moral organ. Their own view is that current work in neuropsychology does not support the idea of a dedicated, domain-specific moral organ, and if anything, supports the alternative, domain-general view. Although the existing data may be revealing with respect to moral cognition, they don’t yet illuminate the linguistic analogy. Consider the existing work on psychopaths and patients with VMPC damage. Neither group shows selective damage in the moral sphere, which Mallon and Prinz take to be strong evidence against a dedicated moral faculty. But, for both theoretical and methodological reasons, we disagree. Many of the current tests of patients thought to have deficits in the moral sphere have not addressed

the issues raised by the linguistic analogy. For example, the published work on prefrontal lobe patients is based on moral reasoning tasks, in particular, Kohlberg's battery of tests, which measure moral maturity based on the content of justifications rather than the nature of the judgments. Because of their emphasis on conscious reasoning, these measures aren't particularly revealing with respect to intuitive judgments, such as those tapped by the dilemmas featured in our web-based experiments, recent functional neuroimaging studies (Greene, 2004, #47), and the new collaborative work reviewed above on moral judgment in individuals with VMPC damage. Further, all of the tests administered to psychopaths, that are morally relevant, focus on the conventional-moral distinction, in which subjects distinguish between unambiguous conventional transgressions and unambiguous moral transgressions, but never between right and wrong. Furthermore, such tests have not included moral dilemmas where there are no obvious norms to adjudicate between different choices, where both choices lead to harm, for example.

At a theoretical level, we are open to the possibility that even the domain-specific components of the moral faculty may be divisible into discrete units. Indeed, some of the evidence we have presented in this discussion point to just such a multi-system model. Some moral principles appear to be available to conscious reflection, while others do not. Patients with emotional deficits show abnormal moral judgments on some

dilemmas, but not others. We argue that such evidence, far from delivering a blow to the linguistic analogy, is in fact an encouraging sign of the type of refinements to models of moral judgment that have been occurring for decades in the research on language. The language faculty includes subsystems for phonology, morphology, semantics and syntax, and even these subsystems can be further divided. For example, recent work on dysgraphic patients (Miceli, Capasso, Banvegna, & Caramazza, 2004) has revealed individuals with deficits in the representation of vowels, others for consonants, highlighting the distinctive neural foundations for these linguistically specific distinctions.

Let us end as we started with a comment by Oscar Wilde: "I choose my friends for their good looks, my acquaintances for their good characters, and my enemies for their good intellects." We couldn't be more pleased to have such excellent "enemies" as Prinz and Mallon in an area of research that is fueled with excitement, passion, and hope for fundamental discoveries about the nature of moral thought and action. As we have tried to clarify, by drawing on analogy to language, we raise new questions about the nature of our moral psychology. In particular, we force empirically minded researchers interested in the nature of our moral judgments to tackle five distinctive questions: 1.) What are the principles that characterize the mature state of moral competence? 2.) How is this moral knowledge acquired? 3.) How does our moral competence

interface with those systems entailed in performance? 4.) How did our moral competence evolve? 5.) To what extent are the mechanisms underlying our moral competence domain-specific? We are nowhere near any resolution on any of these questions, and thus nowhere near a thumbs up or down for the linguistic analogy. With these questions in mind, however, and with answers forthcoming, we can be confident that our understanding of moral knowledge will rapidly deepen.