



Topics in Cognitive Science 2 (2010) 486–510
Copyright © 2009 Cognitive Science Society, Inc. All rights reserved.
ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2009.01064.x

The Linguistic Analogy: Motivations, Results, and Speculations

Susan Dwyer,^a Bryce Huebner,^b Marc D. Hauser^c

^a*Department of Philosophy, University of Maryland*

^b*Department of Philosophy, Georgetown University*

^c*Departments of Psychology, Human Evolutionary Biology, and Organismic & Evolutionary Biology, Harvard University*

Received 1 April 2009; received in revised form 3 July 2009; accepted 20 August 2009

Abstract

Inspired by the success of generative linguistics and transformational grammar, proponents of the linguistic analogy (LA) in moral psychology hypothesize that careful attention to folk-moral judgments is likely to reveal a small set of implicit rules and structures responsible for the ubiquitous and apparently unbounded capacity for making moral judgments. As a theoretical hypothesis, LA thus requires a rich description of the computational structures that underlie mature moral judgments, an account of the acquisition and development of these structures, and an analysis of those components of the moral system that are uniquely human and uniquely moral. In this paper we present the theoretical motivations for adopting LA in the study of moral cognition: (a) the distinction between competence and performance, (b) poverty of stimulus considerations, and (c) adopting the computational level as the proper level of analysis for the empirical study of moral judgment. With these motivations in hand, we review recent empirical findings that have been inspired by LA and which provide evidence for at least two predictions of LA: (a) the computational processes responsible for folk-moral judgment operate over structured representations of actions and events, as well as coding for features of agency and outcomes; and (b) folk-moral judgments are the output of a dedicated moral faculty and are largely immune to the effects of context. In addition, we highlight the complexity of the interfaces between the moral faculty and other cognitive systems external to it (e.g., number systems). We conclude by reviewing the potential utility of the theoretical and empirical tools of LA for future research in moral psychology.

Keywords: Linguistic analogy; Moral judgment; Moral faculty; Moral development

1. Introduction

It has long been recognized that an adequate moral theory must be consistent with facts about human psychology (see, e.g., Bentham, 1780; Flanagan, 1991; Hume, 1789/1888). However, systematic investigations into the cognitive and neurobiological mechanisms that are responsible for moral judgment and morally significant action have only recently begun (Blair, 2007; Greene, 2007; Haidt, 2001; Hauser, 2006; Nichols, 2004; Smetana, 2006). A number of theoretical hypotheses are represented in the emerging literature. We focus on the theoretical and empirical pay-offs of taking seriously the strong analogy between the cognitive mechanisms that are responsible for our moral capacities and those that are responsible for our linguistic capacities. In calling the analogy “strong,” we intend a commitment to the way of studying the mind and cognitive capacities that has emerged since the beginning of the cognitive revolution over 50 years ago, spearheaded by, among others, generative linguists. In this sense, to adopt the so-called linguistic analogy (LA) is to recognize more than its merely heuristic virtues, embracing a certain substantive, though broad, view of the human mind and cognition.

LA represents an approach to the study of moral cognition inspired by the successes of the cognitive revolution. As part of the study of human cognition generally, it is firmly committed to the importance of the competence–performance distinction, to the interdependence of synchronic and diachronic explanations of cognitive capacities, and to the necessity of adopting the computational level as the appropriate level of analysis. These issues, in turn, raise important lines of empirical inquiry, many of which have never been explored in the rich history of research in moral psychology but are beginning to emerge as part of a focused effort to explore the nature of moral knowledge.

Drawing inspiration from the empirical and theoretical methodologies of generative linguistics, LA seeks a description of the mental structures and computations that implement the ubiquitous and apparently unbounded human capacity for making moral judgments. In the process, proponents of LA hope to explain how and why every child develops a moral sense, and how this capacity constrains the range of humanly possible moral systems. Following Chomsky (2005), LA adopts a Galilean approach to the complex target of morality. Instead of providing an explanation for morality writ large, LA seeks to unpack this domain of knowledge, laying bare the computational systems and cognitive interfaces that suffice to guide our moral judgments and decisions. In this way, LA attempts to simplify the moral domain to make it more empirically tractable. Finally, LA attempts to fulfill the demands of descriptive and explanatory adequacy (Chomsky, 1965) by offering an explanation of the synchronic production of mature moral judgments as well as a diachronic account of the ontogeny and phylogeny of our moral capacities (Hauser, 2006; Mikhail, 2007).

Although several of the themes of this paper have been reported and discussed at length elsewhere (e.g., Dwyer, 1999, 2009; Hauser, 2006; Mikhail, 2010), our aim here is to provide a succinct overview of LA and to present a set of new empirical results that establish it as a highly plausible and fruitful approach to the study of moral psychology. We begin by laying out the theoretical motivations that underpin LA (Section 2). We then review some

recent empirical findings that have been inspired by LA and that provide direct empirical support for LA (Section 3). Finally, we conclude by considering the role that LA could play in future theoretical and empirical approaches to the study of moral cognition (Section 4). Although we acknowledge that there are other ways to think about the source and content of our moral psychology (e.g., Haidt, 2001; Kohlberg, 1981/1984; Nichols, 2004; Prinz, 2006; Sripada & Stich, 2006; Turiel, 1983), our primary goal is to demonstrate that LA raises new questions that have failed to be acknowledged by alternative theories. Along the way, we also aim to make clear how certain kinds of empirical findings, insofar as they falsify particular predictions of LA, would undermine the desirability of pursuing LA, or at least entail a significant modification of the strong version of the analogy with language.

2. Motivations

John Rawls (1971) suggested that intuitive judgments about justice are systematic. This proposal provided an initial theoretical motivation for LA, suggesting that humans may possess tacit knowledge of a small set of fundamental moral principles, and that this knowledge is analogous in important respects to the implicit grasp of the fundamental principles of grammar that underwrite our linguistic judgments.¹ Rawls' suggestion has often been misinterpreted as a prescriptive claim (see Mikhail, 2010), but it has also been at the heart of 30 years of debate among philosophers, who have deployed moral thought experiments intended to elicit our intuitive moral judgments. A consistent, but still striking finding is that the principles whose acceptance would explain the patterns that we find in our intuitive judgments are rarely consciously accessible to those making them, and the philosophers who have designed such thought experiments often struggle to explain why some scenarios seem to give rise to the moral judgments that they do. Recent attempts to develop the details of the analogy between moral cognition and linguistic cognition as a guiding hypothesis for the study of moral psychology (Dwyer, 1999, 2006, 2009; Harman, 1999; Hauser, 2006; Jackendoff, 2007; Mikhail, 2007; Stich, 1972) suggest that there are at least three themes, central to generative linguistics, that can profitably be extended to and explored within moral psychology: (a) a distinction between competence and performance; (b) poverty of stimulus considerations; and (c) adopting the computational level as the proper level of empirical analysis.²

2.1. Competence and performance

Linguists have long recognized that linguistic *performance* includes a variety of complexities that stand in the way of directly modeling the cognitive principles that are responsible for language production and comprehension. Speakers utilize incomplete sentences and punctuate their speech with neutral vowel sounds like “um” and “er.” More importantly, the capacities to speak, understand, and acquire a natural language do not require facility with these complexities. Speakers of natural languages are able to understand and produce a

discrete infinity of expressions. Thus, this ability must be grounded in a set of generative rules—a *competence*—that cannot merely be “read off” its output.

The importance of the competence-performance distinction is also captured and reinforced by contrasting the generative grammar of a language (or I-language) with the folk idea of “a” language (e.g., English, French, etc.). Languages (in the folk sense, or E-languages) are local, historically contingent phenomena that are grounded in economic, social, political, and cultural conventions. While the mind-external features of these languages are clearly grist for the sociolinguist’s mill, they are not the only plausible explananda for cognitive science. After all, descriptions of E-languages cannot explain the generative principles that yield our relatively unbounded capacities for linguistic expression and comprehension. Although such data do provide evidence for the necessary existence of an underlying linguistic competence, such descriptive corpora tell us only what *is*; they cannot begin to provide explanations of what *is not*, and why it is not. The set of humanly possible languages is not identical to the set of logically possible E-languages, and this raises an important question: What is the space of humanly possible languages? Answering this question calls for a deep understanding of human linguistic competence, which in turn requires discovering the constraints and boundaries on humanly possible languages. This cannot be achieved by a mere cataloguing of languages; it requires experimentation to assess what is acquirable, what is not, and why. Note, however, that this does not negate the importance of linguistic performance and the E-languages that are generated. Rather, it lays out a methodological strategy that distinguishes between mind-internal processes (both language-specific and more domain-general) and the systems that externalize the outputs of those processes (e.g., articulation and pragmatic mechanisms that are sensitive to contingent features of the speaker’s physical and social environment). Such a strategy thus distinguishes between linguistic knowledge (competence) and how that knowledge is put to use (performance).

Proponents of LA also distinguish moral *competence* from moral *performance*. Thus, LA highlights the importance of adopting a parallel distinction between I-morality and E-morality. E-moralities are historically contingent phenomena that are the result of economic, social, political, and cultural conventions (including a hodgepodge of rituals, customs, practices, institutions, and behaviors); I-morality, in contrast, includes the generative principles that are required if one is to acquire a morality at all. Although descriptive accounts of folk-moralities are of great interest from the standpoint of many psychologists and anthropologists, we contend that they will always be explanatorily incomplete. Moral competence requires the capacity to produce and understand an unbounded range of morally significant judgments and actions. We are frequently faced with previously unexamined cases where moral judgments must be made. As the ease of eliciting moral judgments by way of thought experiments suggests, we have little difficulty arriving at nearly immediate moral judgments even in unfamiliar cases. Thus, moral psychology cannot rest content with a mere catalog of extant moral judgments. After all, such extensionally characterized “moralities” cannot *explain* how we make moral judgments in unfamiliar cases; and this is the central issue to be addressed by cognitive scientists interested in the operation of our moral psychology.³ Moral psychologists, like linguists, owe us an account of why human moral systems take the forms they do, *and not others*. Again, while cataloguing moral

systems is of interest and a critical first step towards delimiting a set of empirically tractable phenomena, it remains only a first step insofar as it fails to engage the possibility that there are other moral systems that *could* be acquired, and yet others that could *never* be acquired (Hauser, 2009).

On the basis of such considerations, proponents of LA suggest that humans are equipped with a moral faculty (FM) that allows us to acquire a morality. FM shapes our capacity for making moral judgments, and in this sense, the computational principles instantiated in FM constrain the range of moralities that can be acquired and deployed in moral reasoning. Paralleling the case of language, this approach positively demands an experimental approach to moral psychology—one that abstracts away from the moral performance manifest in debating the ethics of abortion or in acting in accordance with one's moral judgments, to uncover the underlying cognitive mechanisms at play in the evaluation of the moral status of various actions. As we argue more directly below, studying the FM as an analog to the language faculty (FL) opens up a host of theoretical questions central to contemporary research in the cognitive sciences. Specifically, it calls for an account of which processes are unique to the FM and which nonmoral processes are required for making the sorts of moral judgments that we do. LA thus requires us to ask about the extent to which the cognitive processes involved in theory of mind, or action perception and segmentation, interface with FM, as well as the extent to which FM is a domain-specific system, and the extent to which it is informationally encapsulated. Additionally, LA raises new strategies for investigating the ontogenetic development of moral competence.

2.2. Poverty of the stimulus

Questions about the ontogeny of our moral judgments lead directly to an important fact about human moral psychology. Every typically developing child, in every part of the world, acquires the capacity to understand and produce a discrete infinity of expressions in one or more natural languages; yet no matter how atypical they are, chimpanzees do not, and neither do dogs, who are exposed to much of the same input as are children. Both empirical evidence (e.g., Crain & Nakayama, 1987; Crain & Thornton, 2006; Marcus, 1993) and conceptual arguments (at least back to Chomsky, 1959) suggest that the acquisition of a first language cannot be exhaustively explained in terms of domain-general learning mechanisms (Crain & Pietroski, 2001; Laurence & Margolis, 2001). The primary linguistic input that is available to children at various stages of development is impoverished relative to their linguistic capacities; so poverty of the stimulus considerations suggest that children must rely on domain-specific computations that shape the space of humanly acquirable languages.

Proponents of LA contend that our moral capacities are species-typical in the same sense as language is (Dwyer, 1999, 2006, 2009; Hauser, 2006; Mikhail, 2010).⁴ Again, although dogs adhere to some of the rules that we set for them, and appear to have a sense of fairness that may represent a derived cognitive trait (Range, Horn, Viranyi, & Huber, 2009), they do not appear to have an intuitive sense of moral rightness and wrongness of the sort that every typically developing child eventually acquires, one that includes a sense of “*oughtness*.” From the standpoint of LA, it becomes clear that any systematic explanation of human

moral competence must be grounded in a clear sense of the capacities that children possess at various points in development. This is one respect in which moral psychology lags behind linguistics, which has been able to amass a plethora of evidence concerning what children know about their language and how they come to acquire that knowledge.

This is not, of course, to deny that philosophers, legal theorists, and theologians have—for quite some time—been interested in providing an account of the principles that underlie moral judgment. Indeed, the systematic investigations into the nature of moral rules that we find in Cicero, Aquinas, Grotius, Kant, and Sidgwick were attempts to do just this. Moreover, as Mikhail (2009) has recently been at pains to argue, the more recent attempt to systematize common legal rules in the Model Penal Code and the Restatement of Torts, among others, is a contemporary attempt to offer a systematic investigation of the underlying principles that are responsible for the structure of mature moral judgments. However, whatever else may be said about our knowledge of *mature* moral cognition, at present, we still do not have a comprehensive descriptive account of the developmental trajectory of moral competence. With this in mind, a few remarks are in order by way of motivating some new questions about the psychology of moral development.

First, the target explanandum in moral psychology is neither the presence of morally good behavior (cf. Dunn, 1999; Eisenberg et al., 1999; Hoffman, 1983) nor the ability to assert moral rules (e.g., “Lying is bad”). Such behaviors and assertions, though of great interest, have wildly different causes and, by themselves, suggest no account of their production. The psychological phenomenon that calls for explanation is how young children have any moral capacities at all. By 39 months, children distinguish between moral and nonmoral transgressions (Helwig & Turiel, 2002; Smetana & Braeges, 1990; Turiel, 1983). At approximately the same age, children recognize the special force of permission rules (expressed as conditionals) and, on the basis of this, seem able to attribute intentions to actors in ways that ground responsibility ascriptions (Harris & Núñez, 1996; Núñez & Harris, 1998). Furthermore, as Leslie, Knobe, and Cohen (2006) report, children manifest the so-called “side-effect effect” (Knobe, 2003): They construe unintended but foreseen side effects of an action as intended when the outcome of the action is bad, but they do not do so when the outcome is good. In addition to these rather subtle normative capacities, there are further important questions about how children acquire the specific moral signature of their cultures and bring their moral cognition to bear on their behavior.

Second, although children do receive some moral instruction, it is not clear how this instruction could allow them to recover moral rules (Dwyer, 1999). The world does not come with the required labels for distinguishing moral violations from violations of local norms of etiquette, and the explicit rules handed down by religion do not map onto the psychological distinctions that are in play in either various strands of moral philosophy or in the intuitive judgments that adults offer in both familiar and unfamiliar moral situations. Moreover, when children are corrected, it is typically by way of post hoc evaluations (e.g., “You should not have hit your brother”), and such remarks are likely to be too specific and too context dependent to provide a foundation for the sophisticated moral rules that we find in children’s judgments about right and wrong. Parents do, of course, offer universal imperatives (e.g., “Always keep your promises!”), but a key question remains regarding how

children work out when exceptions to these explicit moral generalizations apply and when they do not. Moreover, it is unclear what can be inferred from these imperatives. What does a child learn when she or he hears “always” or “never” in relation to a norm transgression? These logical operators appear in moral and nonmoral contexts, but how do they interface with morality-specific contents? Furthermore, it seems that nonmoral (i.e., conventional) transgressions are corrected far more frequently than moral transgressions (Nucci, 2001; Smetana, 1989). This leaves open a puzzle: If correction is common for conventional transgressions, but infrequent for moral ones, how is a child to learn moral rules from standard reinforcement? Finally, most children do not commit or witness acts that would warrant correction or instruction of a kind that could account for their broad capacity to make moral judgments, especially when presented with unfamiliar, and often apocryphal cases. Unfortunately, such questions have yet to be directly probed from the standpoint of moral psychology—so we do not have clear answers regarding the precise developmental trajectory of our moral capacities. This is due, in part, to the fact that such questions only arise, and take on their significance, once it is recognized that there is a difference between moral competence and moral performance, and once the generative capacity for moral expression is recognized and explored as an empirical problem.

We acknowledge that, at present, it cannot conclusively be established that the moral input is impoverished relative to the child’s moral capacities. To establish that this is the case, we would need to have a richer account of:

1. The mature state of moral knowledge (which will benefit greatly from targeted analysis of mature responses to moral dilemmas discussed below, as well as investigations of commonalities and differences in legal codes and explicit moral theories);
2. The precise state of moral cognition at various stages of moral development;
3. The processes that underpin the acquisition of the capacity for moral judgment; and,
4. The precise connection, if any, between correction and patterns of change in moral judgment.

Without targeted investigations into these questions in the moral domain, we suggest that it is a live possibility that moral competence develops through a process analogous to language acquisition. The existing data suggest the following view of moral development: The human mind includes a biological mechanism that provides a limited range of possible moral systems, and on this basis, the environment selects among the cultural options on offer to acquire a particular morality. However, if future data reveal that children can and do simply extract moral principles from their environments, relying on nothing more than the pattern of corrections and affirmations that they encounter early in life, then poverty of the (moral) stimulus considerations will have to be abandoned. Similarly, if future studies reveal that there is no stable developmental trajectory for moral cognition, paralleling linguistic growth, or that the moral principles that children rely upon for moral judgments are intimately tied to the cultural environments in which they have been raised, then the claim that moral competence is richer than the available input will also seem less plausible.

Regardless of whether a particular poverty of the moral stimulus argument succeeds, we contend that researchers cannot determine the extent to which experience alone is sufficient to explain specifically moral development without first isolating the components of moral competence and considering how these interact with nonmoral aspects of the human mind. Only with such descriptively adequate understanding in hand can we tackle the problem of moral acquisition. This is a general methodological point that has typically been overlooked but that is made vivid by adopting LA.

2.3. *How to study competence*

The most successful method for recovering the computational principles of the FL has been driven by the elicitation of acceptability judgments, either from native speakers with no formal training in linguistics or from linguists themselves. Competent speakers are asked (a) to judge whether particular strings are acceptable in their language; (b) to rank strings in order of acceptability; and (c) to identify possible and impossible meanings of particular strings. For example:

1. *She liked pictures of each other
is judged unacceptable by competent English speakers (as indicated by the asterisk); and
2. Which film did you wonder which critic liked?
is ranked as more acceptable than
3. Which critic did you wonder which film liked?
and
4. The architect called the builder from Boston

String 4 will be judged to mean either that the architect made the call from Boston, or that the architect called a builder who was from Boston. However, no competent speaker will take it to mean that the architect was from Boston.

Linguists use acceptability judgments to test candidate grammatical rules that they believe to be instantiated in the computational architecture of FL. Ordinary speakers' acceptability judgments are remarkably convergent and confident. However, those same speakers are typically unable to explain the computational principles that are responsible for their judgments of grammaticality.

As we noted at the outset, eliciting folk-moral judgments (as compared to explicitly theoretical moral judgments derived from consciously entertained moral principles) has long been a mainstay of moral philosophy. Philosophers have often constructed thought experiments to evoke moral judgments in support of a variety of moral theories (e.g., consequentialist and nonconsequentialist theories), principles (e.g., the Doctrine of Double Effect),⁵ and morally significant distinctions (e.g., between actions and omissions) (e.g., Foot, 1967; Quinn, 1989; Thomson, 1985).⁶ Moral psychologists, in contrast, evoke folk-moral judgments in the service of uncovering the cognitive machinery that is deployed in moral cognition, and to identify the features of actions and agents that are salient to moral cognition. In much the same way that individuals respond to the grammaticality of a sentence, individuals appear to

spontaneously and confidently offer moral judgments in response to moral dilemmas. Proponents of LA contend that these intuitive moral judgments do not typically express reflectively held normative principles (Cushman, Young, & Hauser, 2006; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Hauser, Young, & Cushman, 2008; Mikhail, 2007) but instead are driven by reflexive computations that can be uncovered by systematically manipulating the important features of moral scenarios and testing various target hypotheses about the representations over which moral cognition operates.⁷ Paralleling our comments above, LA would be at least partially defeated, then, if moral intuitions were typically and readily revisable. In other words, sustaining LA depends in part on establishing that the computations responsible for folk-moral judgments are distinct from and, at least relatively, immune to the results of conscious moral reflection on new, potentially morally relevant data.

Keeping these general motivations for pursuing LA in mind, we now turn to a set of experimental results that have been derived from LA, and that provide empirical support for this perspective on moral cognition. These results begin to provide more detailed answers to deep questions about how moral cognition works, thereby demonstrating the value of LA as an approach to moral psychology. Specifically, the work we discuss bears on the following central predictions of LA:

1. The computational processes responsible for folk-moral judgment operate over structured representations of actions and events, as well as coding for features of agency and outcomes.
2. Folk-moral judgments are the output of a dedicated FM and are largely immune to the effects of context.

In addition, this work underscores the complexity of questions about how the FM interfaces with other cognitive capacities.

3. Tests of the linguistic analogy

We read newspapers and novels, watch news broadcasts, and listen to public radio. Each of these experiences presents us with morally relevant and morally irrelevant input. Yet spontaneously, and often without reflection or the ability to deliver a coherent explanatory justification, we evaluate the moral status of the actions embedded in these vignettes of life, and we do so without engaging in further morally significant actions. To explain this facet of our moral psychology, a viable account of moral judgment owes a description of the computational structures that implement the analysis and evaluation of actions prior to the production of moral judgments (Hauser, 2006; Mikhail, 2007, 2008). To make such judgments, we must represent agents, actions, and outcomes, as well as the relations between these. Moreover, we must be able to distinguish various action types: Crooking one's finger when it is on the trigger of a loaded gun is quite different than crooking one's finger in mid air. The key question suggested by LA, then, concerns the computational and representational capacities that implement moral judgment.

According to LA, the FM is a domain-specific system; however, the complexities of our moral psychology also require that this system interface with a variety of other computational systems (see Hauser & Young, 2008; Sinnott-Armstrong, Mallon, McCoy, & Hull, 2008; Waldman & Dietrich, 2007; Young & Saxe, 2008). However, at present it is unclear which systems interface with moral cognition. To make the import of this latter claim clear, consider the recent debates over the role of emotion in our moral psychology. Huebner, Dwyer, and Hauser (2009) argue that currently available data are both insufficient to demonstrate that emotion is causally implicated in the production of moral judgments, and insufficient to establish that emotion is critically required for the development of the capacity for making moral judgments. This does not mean that emotion is unimportant to our moral psychology; nor does it mean that there are no important interfaces between moral cognition and emotion. Indeed, one view consistent with LA is that various emotional processes are likely to be brought on line to motivate morally significant actions, antecedent to the computations required for making moral judgments. Thus, proponents of LA hypothesize that emotional representations play an important role in guiding moral performance, although they are not part of our FM, narrowly construed.⁸

By adopting LA, empirical inquiries are focused upon the cognitive capacities that allow us to discriminate morally significant aspects of our environment. In the following text, we discuss three ways in which LA has allowed us to gain traction on the computations that implement our capacity for making moral judgments. To be clear, we are not saying that LA is the only route into these questions nor that LA provides the best explanatory account of all of the results in moral psychology to date. Rather, our point is that LA demands answers to questions that have not previously been addressed, that interesting experiments have been motivated by LA, and that the results of these experiments provide some confirmation of the hypothesis that we possess a FM that operates, in the context of the whole human mind, in a way that is analogous to the operation of our FL.

3.1. Parsing harm

A number of studies relying on *Trolley-type* dilemmas have demonstrated that folk-morality cannot be exhaustively characterized in either consequentialist or nonconsequentialist terms (Cushman et al., 2006; Greene, Cushman, Stewart, Nystrom, & Cohen, 2009; Mikhail, 2007). Participants in moral psychology experiments typically judge it permissible to sacrifice one person to save the lives of five others when this requires redirecting a trolley, but impermissible when this requires pushing a nearby person in front of the trolley. One explanation of this difference relies on the claim that our folk-morality is sensitive to the manner in which a morally significant outcome is brought about. In particular, from the standpoint of folk-morality it seems to be (at least *prima facie*) impermissible to harm one person without her consent as a means to achieving some morally preferable end (Cushman et al., 2006; Hauser et al., 2007, 2008; Mikhail, 2007, 2008). Another explanation of this difference is that folk-morality conforms to the Principle of Double Effect (see note 5). Building on recent neuroimaging data, Greene and colleagues (Greene, 2007; Greene & Haidt, 2002; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Greene et al., 2009)

have suggested a third possible explanation of this difference: Folk-morality distinguishes between personal harms, which require a direct physical action to bring about a morally significant outcome, and impersonal harms, which require instead a causal intermediary (such as flipping a switch) to bring about that outcome.

Although each of these factors may play an important role in our moral psychology, prior experiments have often introduced another, confounding factor that has typically been overlooked: The *degree* to which the person who is harmfully used to bring about a greater good is made worse off. In the Footbridge case, the man who must be pushed in front of a runaway trolley to save five people would be fatally harmed in a way that would otherwise be avoided. But what if this man was already doomed? Bernard Williams (Smart & Williams, 1973) considers a case in which Jim stumbles upon a small village where a military leader is about to execute 20 natives. The military leader tells Jim that if he shoots one of the Indians, he will set the remaining 19 free—but death is inevitable for all 20 if he does nothing. Hence, even if Jim shoots one person, he will make no one worse off. Although he is a thoroughgoing critic of the Utilitarian view, Williams nonetheless concedes that it would be morally right, though tragically so, to kill the one person. The point to note here is that shooting the one person constitutes a Pareto improvement (Pareto, 1906/1971): It makes the outcome better for 19 people and no one is made worse off.

In a recent experiment, Moore, Clark, and Kane (2008) found that subjects judged those actions that lead to inevitable deaths to be more permissible than actions that introduced the harm of death to an otherwise nonendangered person. Building on these data, B. Huebner, M. D. Hauser, and P. Pettit (unpublished data) examined the interaction between the directness and the inevitability of harm in Pareto and non-Pareto scenarios. Across a variety of contexts, they found that folk-moral judgments were sensitive to considerations of Pareto improvement. Specifically, regardless of the source of an ongoing threat, people judged that it was permissible to use someone as a means to a greater good in those cases where doing so does not make that person worse off. However, Huebner et al. also found that folk-moral judgments about harmfully using a person as a means to some greater good are sensitive to the source of the ongoing threat. In the familiar trolley car scenarios that rely on a mechanical threat, for example, participants judged that actions involving direct physical contact were morally worse than those involving a causal intermediary: Throwing a rock at someone to make them scream and act as an alarm call to save five people was seen as more permissible than pushing a person to the ground to achieve the same end. However, although considerations of Pareto improvement continued to operate across sources of harm, considerations of direct physical contact were attenuated where the source of the ongoing threat was an intentional agent (as in Williams' Jim case) or a nonmechanical threat (e.g., the fire in a burning house). This shows that considerations of Pareto efficiency interact in important ways with the nature or source of the impending threat, highlighting again the relevance of considering the interfaces between different representational systems.

According to LA, intuitive moral judgments are implemented by computations that operate over a set of abstract, content-independent, principles or distinctions. Previous empirical studies (Cushman et al., 2006; see also Moore et al., 2008), building on a rich philosophical literature (e.g., Kamm, 2007; Thomson, 1985) provide evidence for three such distinctions:

(a) harm caused by action is worse than an equivalent harm caused by an omission; (b) harms caused as a means to some greater good are worse than equivalent harms caused as a foreseen side-effect of an action; (c) harms that rely on physical contact are worse than equivalent harms that are brought about by a nonhuman causal intermediary. The data reported by Moore et al. (2008) and B. Huebner, M. D. Hauser, and P. Petit (unpublished data) strongly suggest a fourth fundamental distinction at play in our moral judgments: Pareto improvement—intentionally harming a person as a means to the greater good—is more permissible if the person harmed is not made worse off by the harmful act.

In line with the presence of this additional distinction, these data seem to suggest that when a person is already doomed, actions that are initially treated as bringing about harm may sometimes be recoded, leading the FM to interpret the relevant action as not harmful. In cases where there is a Pareto improvement, the initial aversion to treating a person as a means to a greater good can sometimes be overridden by a computation that *recodes this initial harm*, treating it instead as a passing physical discomfort foisted upon a person who no longer has interests that can be violated or set back (Feinberg, 1984). However, as Huebner et al. found that Pareto-considerations interacted differently with considerations of direct physical contact depending on the context in which the harm occurred, it is also important to note that the significance of different principles or distinctions for our moral judgments is likely to depend on the structure of the moral scenario, including facts about which other distinctions or factors (e.g., whether the threat comes from a human agent, a mechanical object, or a natural danger) are in play within a given moral scenario.

Such data highlight some of the complexities involved in our moral competence. Although moral judgments display a pattern suggestive of principled distinctions (across individuals), and independent of many matters of content, the fact that various factors interact in important ways with one another has yet to be fully appreciated. Although researchers following theoretical perspectives other than LA have also explored how different factors such as contact, intention, and means-based harms guide moral judgments (Bartels, 2008; Greene et al., 2009; Moore et al., 2008; Royzman & Baron, 2002; Waldman & Dietrich, 2007), we suggest that LA provides a unique set of testable hypotheses regarding the operative force of these factors. In particular, if factors such as means versus side effect and Pareto-efficiency are part of moral competence, in the way that computations such as “merge” and “copy” are part of linguistic competence, then not only will these computations be operative independently of particular content, but they will be unconsciously operative, inaccessible to folk intuition, and, even when made explicit for familiar cases, will play no role in guiding judgment in novel and unfamiliar cases. Even where reflective judgments lead to changes in the pattern of expressed judgments for a particular class of cases, these changes will be localized and will not have ramifications for moral judgment more broadly. Thus, for example, telling subjects that a key difference between the trolley bystander case and the footbridge case lies in the distinction between means and side effects should make no difference in future judgments involving structurally similar but unfamiliar cases. Further, if some of these distinctions are specific to the moral domain, then we should find patients with selective deficits, imaging work that reveals selective patterns of activation, and results from TMS (transcranial magnetic stimulation) studies where suppressing

activity in the critical circuitry causes selective elimination of the target distinction in moral evaluation.

3.2. *The role of calculation in the comparison of harms*

Relying on these different considerations regarding our moral competence sets up a fascinating set of empirical studies designed to assess not only which of these factors are specific to the moral domain, but when, in the moral computation, different factors are triggered, and how the mind/brain resolves conflicts that may arise between different considerations. One area in which such considerations are especially pronounced is in questions about the sensitivity of our moral judgments to the maximization of welfare. In moral judgment, is *more* always *better*? It seems plausible that folk-moral judgment would be sensitive to numerical considerations, treating actions that save five lives as morally better than those that save just one life, and actions saving 500 lives as better than actions saving 100 lives.

The key issue raised at this point concerns the interface between moral cognition and domain-specific systems dedicated to enumeration. At present, there are three broadly recognized capacities for numerical cognition: a core analog magnitude system that approximates the size of large numbers and is limited by Weber fractions (Brannon, 2002; Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004), a core parallel individuation system that operates exactly over small numbers up to 3 or 4 (Carey, 2001; Scholl & Leslie, 1999), and an arithmetic system that operates on the principles of numerical identity and succession (Leslie, Gelman, & Gallistel, 2008). But which of these systems, if any, plays a role in making moral judgments? This question is not merely of theoretical interest; it also highlights the interface between moral cognition and numerical cognition, two presumably distinctive domains with dedicated computations and representations that may interact in interesting ways when we evaluate morally relevant events.

To examine this question, B. Huebner, N. Miller, H. R. Seyedsayamdost, and M. D. Hauser (unpublished data) asked how many lives must be at stake for people to judge it to be permissible—or even obligatory—to intentionally kill a person. They found that people were surprisingly insensitive to many considerations of *quantity* in utilitarian calculations of permissible harm. For example, participants saw no significant difference between redirecting a runaway boxcar, killing one person to save two, and killing 500 hundred people to save 501. More strikingly, when participants were asked how many people would have to be on the initially dangerous track for it to be obligatory to divert a boxcar onto the track where it would kill one person, the modal and dominant response was two lives. This result was stable even in nontrolley scenarios (B. Huebner & M. D. Hauser, unpublished data), suggesting that the effect is robust across moral contexts. In brief, many participants judged it permissible to redirect a lethal threat onto a single person whenever *one more* person could be saved—and many participants even judged that this action was obligatory.

These results set up the hypothesis that judgments of moral permissibility in the context of utilitarian calculations are largely impenetrable to the numerical calculations that are carried out by the two core nonlinguistic systems for enumeration. Instead, it seems that within the context of utilitarian judgments, folk-morality relies on considerations of numerical

identity and numerical succession. Huebner et al. thus offer the *+1 principle* for judgments of permissible harm: As long as the number of lives saved exceeds the number of lives lost by one or more, it is permissible to kill the smaller number of individuals. Building on this relative insensitivity to nonlinguistic numerical considerations, as well as on previous data suggesting strong dissociations between intuitive moral judgments and deliberative ones (Hauser et al., 2007; Hauser et al., 2008), it seems reasonable to hypothesize that the computations underlying folk-moral judgment are informationally encapsulated to a significant degree (Fodor, 1975; Pylyshyn, 1984). However, more data would be required to show that moral computations are indeed informationally encapsulated. One way in which this hypothesis could garner further support is by demonstrating that moral judgment is immune to the well-known cognitive heuristic strategies that are employed in making evaluative judgments under conditions of uncertainty, a condition that arguably obtains in the consideration of moral dilemmas.

3.3. Immunity to context effects

It is well established that people rely on simplifying heuristics in making evaluative judgments (Tversky & Kahneman, 1974). In line with this recognition, a wide range of studies in psychology have shown that apparently deliberative reasoning is often subject to irrelevant effects of context. For example, people often neglect the relevance of situational factors in the explanation of human behavior, instead relying on assumptions about character (Gilbert & Malone, 1995; Jones & Harris, 1967; Ross, 1977). Moreover, evaluations of various experiences are altered by unconscious contrasts with earlier experiences of the same type (Tversky & Griffin, 1991). In short, heuristic strategies underlie many of the decision-making processes that we employ, especially where the correct answer to a question is uncertain. The sorts of decisions that are called for in making moral judgments, especially in making judgments about moral dilemmas, seem to generate precisely this kind of uncertainty. There are no obviously correct (or incorrect) answers about what should be done in the case of a moral dilemma. However, if folk-moral judgments are driven by distinctly moral computations, rather than by domain-general processes of deliberation and reflection, they should not be as reliant on familiar heuristic strategies for reducing uncertainty. Evidence that moral judgment does not rely on the same heuristic strategies that we find in other sorts of evaluative judgment would help to provide support for the claim that some aspects of moral cognition are informationally encapsulated, relying on domain-specific computational principles rather than on domain-general strategies for making judgments in conditions of uncertainty.

B. Huebner and M. D. Hauser (unpublished data) investigated the vulnerability of folk-moral judgments to the effects of order and context by collecting permissibility judgments for three different types of situation, each with 10 different morally salient scenarios. In each condition, half of the participants began with a case that was obviously obligatory (saving five people at no cost) and moved to a clearly forbidden one (killing five people with no benefit); the other half of the participants began with a forbidden action and moved to an obligatory one. If moral judgment is subject to ordering effects or other effects of context,

the responses of subjects with respect to the differently ordered series should be different. In particular, if moral judgments of the initial cases serve as anchors for judgments in ensuing cases, then participants ought to demonstrate consistency with their prior judgments. Specifically, in an experiment such as this, participants' judgments should be affected by whether they began by considering an obviously forbidden or an obviously obligatory act.

Participants were asked to provide judgments about a set of 10 *River Raft* cases in which a dam has broken up-stream and a decision has to be made about whether to divert the water into a nearby drainage canal. In these cases, there was no effect of the order of presentation—suggesting that prior judgment had no impact on subsequent judgments within the context of the experiment. To test the generality of this effect, two similar sets of dilemmas, the first relying on a gas leak in a hospital, the second relying on an out-of-control trolley (“boxcar” in the actual scenarios), were examined. The gas leak cases confirmed the result of the initial experiment, with no order effect on eight cases, and incredibly small effects for the remaining two cases. However, there was an effect of order of presentation for those participants who were asked to make moral judgments about the boxcar/trolley scenarios.

These data suggest that folk-moral judgment—and hence the computations that underwrite it—is predominantly impenetrable to domain-general heuristic strategies. However, it seems that in those cases where people have made prior moral judgments about similar cases, nonmoral computations may be able to intervene in the judgment that is delivered in the context of a moral psychology experiment (we return to this suggestion in the next paragraph). This finding has at least two substantive implications that are worth noting. First, it mitigates the concern that folk-moral judgments are especially likely to be driven by or vulnerable to domain-general heuristics, as Gigerenzer (2008) and Sunstein (2005) suggest, respectively. If things had turned out otherwise, it would have encouraged a skeptical attack on one central idea of LA—namely, that moral judgment is to be explained by reference to principles that characterize the systematic operation of particular computational cognitive processes. However, at least within two of the contexts presented, and the moral scenarios tested, our intuitive moral sense appears immune to reasoning and reflection.

This contrast between *reflective* and *reflexive* judgments connects to a second, substantive point that raises a pressing question for proponents of LA. The significant effect of order of presentation found for participants who made judgments about boxcar/trolley cases suggests that context effects might be tied to a participant's familiarity with a particular moral dilemma. Trolley-like cases have gained a degree of notoriety in public discourse. So familiarity might play the role of a previously endorsed reflective judgment when participants are asked to render reflexive judgments. If this is correct, then some account is owed of the precise relation between the processes involved in reflexive and reflective moral judgment. Methodologically, the possible effects of familiarity should encourage researchers wishing to get at the “unadorned” processes that drive moral judgment to deploy nontrolley scenarios that mimic the kinds of new cases participants are likely to encounter outside the experimental setting.

4. Speculations

The results reported in Section 3 provide evidence for at least two predictions of LA, which we reiterate from Section 2:

1. The computational processes responsible for folk-moral judgment operate over structured representations of actions and events, as well as coding for features of agency and outcomes.
2. Folk-moral judgments are the output of a dedicated FM, and they are largely immune to the effects of context.

In addition, the results discussed hint at the ways in which the FM interfaces with other cognitive systems, generating patterns of judgments that are far from obvious, and certainly not predicted on an a priori basis. While the evidence collected thus far begins to provide support for each of these claims, the theoretical and empirical tools of LA suggest at least four additional issues, which we offer here as invitations for future research.

4.1. Constraints on variation

First, universal grammar (UG) provides an explanation for the fact that all typically developing children acquire language. However, it is often suggested that UG also defines the set of humanly possible languages. Supposing that UG is comprised of a set of principles that are initially underspecified and that mature in particular ways as a result of the child's particular linguistic experiences (which, in turn, are likely tied to the idiosyncrasies of her culture), UG would provide an evolutionarily specified set of developmental constraints on the structure of language. It is often hypothesized that a small set of universal principles account for the universal aspects of natural language, the speed and the trajectory of first-language acquisition. However, the hypothesis that we wish to stress here is that apparent variability across human languages is *constrained* variation that emerges in accordance with these universal principles.

In line with the hypothesis that there are biological mechanisms that limit the range of humanly possible languages, LA raises questions about the nature and scope of the variation that we find across moral communities. LA predicts that there are invariant, universal principles that lie at the core of the FM, perhaps with parameters providing options to create cross-cultural variation. Of course, there is a long way to go in establishing the nature of these parameters, as well as the kinds of childhood experiences that trigger the setting of parameters. However, the question that emerges in light of LA is, "Does the structure of the FM constrain the range of cultural variation in human moral systems?" That is, is there a universal *moral* grammar that defines the set of humanly possible moralities?⁹ One of the most intriguing aspects of LA is that it allows us to take seriously the possibility of moral relativism; and it does so in a way that suggests a psychological mechanism that could underwrite cultural variation while still yielding the perceived universal applicability of local rules. Rather than resting content with the claim that there is moral

variation—something that is impossible to deny—LA requires us to explain *why* there is such variation, how extensive it is or could be, and to address these problems by exploring how our psychology extracts the computational structures that allow us to have variable moral views. The issues in this area are both theoretically and empirically complex, and they cannot be avoided merely by mapping the space of humanly acquired moralities. Rather, the exciting prospect evoked by LA is that there are some things that will never be possible within the domain of human morality.¹⁰ Just as the set of humanly possible languages are likely to be a subset of logically possible languages, we hypothesize that the range of humanly possible moralities is a subset of the logically possible moralities. If there are such constraints on the range of possible moralities, then it is no surprise that folk-moral judgments often agree with the deliverances of moral theory, suggesting that deep facts about our moral cognitive architecture have significant implications for the range of plausible moral theories applicable to creatures like us.

It may turn out, given further empirical investigations into the ways in which moral judgments differ across cultures, that there are no universal, stable moral principles. Strategies for making moral decisions, and for evaluating the status of a morally significant action, may differ so widely across cultures that they resist explanation in terms of a set of universal computational principles, parameterizable or not. If this is the case, a strong analogy between moral computations and linguistic computations is significantly less plausible. In other words, we would be forced to reject LA as an account of our moral judgments, should it be established that human beings, across or within cultures, employ wildly heterogeneous strategies to arrive at utterly diverse patterns of moral judgments. Fortunately, however, the emerging range of data suggests constraints on the scope of possible variation. Specifically, what we know to date suggests that the perceived variation in moral judgments and practices is likely to be constrained in precisely the sense that is predicted by LA (e.g., L. Arbarbanell & M. D. Hauser, unpublished data; Henrich et al., 2001, 2006).

4.2. *Developmental constraints on moral cognition*

LA requires a focus on both mature moral judgments and the development of the capacity to make moral judgments. To date, much of the research that has been conducted in light of LA has targeted only mature moral judgments. However, this rubric raises exciting questions regarding children's moral development, a systematic way of addressing them, as well as the possibility of integrating findings from other areas of developmental cognitive science—for example, theory of mind, agency-detection, “core” knowledge systems (Spelke & Kinzler, 2007)—with results concerning moral judgment (see Leslie, Mallon, & DiCorcia, 2006; Leslie, Knobe et al. 2006). For example, as we noted above, just as adults' assessment of moral character has a downstream effect on intuitions about intentionality and causality (Knobe, 2003), such effects are visible in young children. Such developmental studies will help to determine how morally specific representations and computations, if there are any, interface with more domain-general processes to create the unique signature of our domain of moral knowledge.

The idea of a domain-specific FM is controversial (Mallon, 2008). Acquiring data about the point at which children achieve the cognitive milestones that are required for making moral judgments will allow for the investigation of which processes are developmentally necessary and which developmentally sufficient for moral judgment. LA predicts that there will be a typical developmental pattern for the emergence of the various capacities that are at work in making moral judgments. Moreover, LA predicts that the emergence of moral capacities will not be exhaustively predicted by patterns of moral correction, and in an important sense, the underlying capacity will be immune to attempts at correction. However, if empirical inquiries into the development of moral cognition show that this is not the case, LA will seem far less plausible as an account of moral cognition.

4.3. *The evolution of moral cognition*

We, and other proponents of LA, claim that one of the virtues of this perspective is that it sharpens evolutionary questions concerning morality. Specifically, because of its emphasis on identifying the computations underlying moral judgment, LA forces researchers to be much clearer about the extent to which the FM relies on domain-general cognitive processes and to what extent it embodies morally specific processes. Thus, just as we speculate that pursuing LA will answer questions about what is unique to the moral system, we anticipate that doing so will also help to reveal which aspects of the FM are unique to our species. While LA assumes that the FM is a biologically determined cognitive feature of *Homo sapiens*, it is crucial to note that some nonhuman animals, especially the higher primates, live in social groups characterized by dominance hierarchies, and evince behaviors—like cooperation, reconciliation, and the punishment of transgressors—that suggest that they too are endowed with some degree of normative sensitivity. For example, in studies of monkeys and apes, there is evidence that individuals pay attention to both the means and outcomes of events. Specifically, chimpanzees show signs of frustration when an experimenter intentionally presents and then withholds food in an act of teasing, but they are unaffected by a clumsy experimenter who fails to give them food (Call, Hare, Carpenter, & Tomasello, 2004). Further, tamarins are more likely to cooperate with an individual who intentionally gives food than with an individual who delivers food as a byproduct of otherwise selfish behavior (Hauser, Chen, Chen, & Chuang, 2003). This latter capacity provides a critical foundation for evaluating the ways in which our conceptions of justice and our dispositions for engaging in reciprocal relationships have evolutionary precursors. Most importantly, though, the thought is that LA will allow us to identify what the human mind/brain adds to basic social primate cognition to give us what we call morality.

4.4. *The implementation of moral cognition*

The fact that LA proceeds in terms of systematically studying the computations underlying the components of moral competence that have begun to be identified on the basis of a developing descriptive account of that competence, critically dovetails with work in the new but large and active field of the neuroscience of morality (Anderson, Bechara,

Damasio, Tranel, & Damasio, 1999; Blair, 2008; Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Ciaramelli, Muccioli, Làdavas, & di Pellegrino, 2007; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Koenigs et al., 2007; Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2007; Moll et al., 2002). Attempts to “map the moral brain”—regardless of the particular capacities they target—face two challenges that confront the cognitive neurosciences quite generally. The Granularity Mismatch Problem and the Ontological Incommensurability Problem (Poeppel & Embick, 2005; Poeppel & Monahan, 2008) remind us that the principles and distinctions used in cognitive science are of a different “grain” than those used in neuroscience. For example, syllables and morphemes do not *obviously* map onto the primitive elements of neuroscience, for example, neurons and cell assemblies. While there may indeed be discrete cortical networks that implement each syllable and morpheme, this is something that we are a long way from being able to demonstrate on the basis of current hemodynamic and electrophysiological technologies for investigating neural activity in humans. In addition, the computations that are posited as operating over the primitive elements, as they are identified in any cognitive science (e.g., in linguistics, concatenation), are not commensurable with our current accounts of neural processes.

At the end of the day, we would like to understand how the computational processes operative in the production of moral judgments are implemented at the neural level. However, a crucial first step toward that understanding is to get the elements and processes posited in moral psychology into the right “shape” so that research can proceed to identify how the brain makes moral judgment possible. To reiterate a major theme of this paper, we encourage researchers to look deeper below the surface of moral judgments and to formulate and test hypotheses about the computations that underwrite such judgments. With these principles in mind, it may be possible to construct a computational model of mature moral cognition, which in turn can lead to more systematic investigations into the implementation of moral cognition in mature humans. Furthermore, pursuing LA makes clear that we will need a vocabulary in which to state these hypotheses. The central notions of “judgment,” “cognition,” and “emotion,” for example, are both abstract and coarse grained. It is impossible to say with any confidence at all where in the brain *judgment, cognition, and emotion* occur. Neuroimaging studies can, of course, reveal which areas of the brain are activated during belief attribution, computations of utilitarian outcomes, certain emotional experiences, and so on. However, these techniques cannot, as yet, help us understand how we represent the complex hierarchical structure of actions, and how we determine whether we are confronting a moral dilemma as opposed to some other socially significant nonmoral dilemma. Similarly, though patient studies can reveal which neural regions are perhaps necessary for particular moral computations, we are still a long way from any detailed neurological account of our moral capacities. As LA begins to offer clearer and more detailed accounts of the computations carried out in making moral judgments, this model will help to target investigations into the implementation of these computations.

5. Conclusion

Adopting LA as a working model for studying moral cognition emphasizes the necessity of being clear about what constitutes empirically tractable explananda for moral psychology; it focuses attention on the computations that underlie moral judgments; and it proposes a variety of causal factors which can be systematically manipulated in order to discover how those computations work. Proponents of LA do not contend that the analogy between language and morality is perfect. Of course, the function and content of moral computations are different from the function and content of linguistic computations. However, LA facilitates the empirical investigation of our moral competence by getting the phenomena into the right shape for systematic and rigorous inquiry. In treating moral competence as a biologically grounded feature of human minds, LA connects directly with human developmental, cross-species, and neuroscientific work, with all of their respective challenges.

Targeted research inspired by LA is in its infancy, barely ahead of its development as a theoretical model of moral competence. But this is as it should be in any nascent scientific domain. We should expect mutual adjustments between theoretical models and experimental methodologies as work progresses. Perhaps most important, LA is a well-motivated approach to the empirical inquiry into human morality that, in putting its hypotheses to the test, permits its own vindication or falsification in the face of the facts.

Notes

1. Rawls is not the first to offer this suggestion; Smith (1817) also adverts to the potential analogy between the rules of grammar and the rules of justice.
2. A proponent of one version of LA, Jackendoff (2007) is hesitant to draw such a direct analogy between linguistic competence and moral competence. Specifically, Jackendoff claims that although grammaticality judgments are merely side effects of being competent to use language for its communicative purpose, moral competence has the explicit purpose of producing moral judgments. On Jackendoff's view, judgment data collected in response to moral dilemmas may not stand in the same evidential relation to moral competence as grammaticality judgments elicited by linguists stand to linguistic competence.
3. Of course, humans talk *about* morality and the metacognitive processes involved in the life of a creature with basic moral capacities of interest, but they are not the most empirically tractable explananda for moral psychology (cf. Dupoux & Jacob, 2007; Dwyer & Hauser, 2008).
4. We make no claims here about the human uniqueness of this capacity. This is an empirical issue whose investigation can be furthered by adopting LA (see Hauser, 2006).
5. The Doctrine of Double Effect holds that an action that is known to have a bad consequence is permissible only if (a) a good consequence is *intended*; (b) the bad consequence is merely a *foreseen* consequence of the intended action; (c) the bad

- consequence is not a *necessary* means to bringing about the good effect; and (d) the good consequence is sufficiently important to justify the foreseen bad consequence.
6. The practice continues (see Kamm, 2007; Thomson, 2008).
 7. This is not to say, however, that there is no convergence between folk-moral judgments and the philosophically informed intuitive judgments of moral philosophers. Indeed, as we note below, some folk-moral judgments are shared by moral theorists whose own favored theories entail the incorrectness of those judgments. In Section 4, we return to this point in order to illustrate some speculations about the intellectual fecundity of pursuing research under the rubric of LA.
 8. It is also important to note that Huebner et al. (2009) do not intend to claim that there could be no evidence that emotional processes are implicated in the production of moral judgments. If there were clear data demonstrating that emotional evaluations were carried out antecedent to the production of moral judgments (either ontogenetically or within the scope of a particular moral judgment task), this would require a revision of LA to accommodate the fact that affective representations play an integral role in the acquisition of moral competence or the capacity for mature moral judgment. However, existing data fail to make this point clear; so, this hypothesis remains a live open option.
 9. These are excellent questions to ask in moral psychology, in spite of the provocative challenges to the Principles and Parameters model emerging in linguistics (e.g., Biberauer, 2008; Hornstein, 2009; Newmeyer, 2005).
 10. The notion of possibility here is broad, encompassing the idea that there might be evaluative frameworks that (a) could not have evolved in the species; (b) are not acquirable by human children; and (c) if acquirable, would not be stable in human societies (Hauser, 2009).

References

- Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience*, 2(11), 1032–1037.
- Aquinas, T. (1963). *Summa theologiae*. London: Blackfriars.
- Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, 108(2), 381–417.
- Bentham, J. (1789). *Introduction to the principles of morals and legislation*. London: T. Payne.
- Biberauer, T. (Ed.) (2008). *The limits of syntactic variation*. Amsterdam: John Benjamins.
- Blair, R. J. R. (2007). The amygdala and ventro medial prefrontal cortex in morality. *Trends in Cognitive Sciences*, 11(9), 387–392.
- Blair, R. J. R. (2008). Fine cuts of empathy and the amygdala: Dissociable deficits in psychopathy and autism. *Quarterly Journal of Experimental Psychology*, 61(1), 157–170.
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, 18(5), 803–817.
- Brannon, E. M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, 83(3), 223–240.

- Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). "Unwilling" versus "unable" chimpanzees' understanding of human intentional action. *Developmental Science*, 7(4), 488–498.
- Carey, S. (2001). Evolutionary and ontogenetic foundations of arithmetic. *Mind & Language*, 16(1), 37–55.
- Chomsky, N. (1959). Review of verbal behavior. *Language*, 35(1), 26–57.
- Chomsky, N. (1965). *Aspects of a theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (2005). *Rules and representations*. New York: Columbia University Press.
- Ciaramelli, E., Muccioli, M., Làdavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2(2), 84–92.
- Cicero, M. (1999). *On the common wealth, and on the laws*. Edited by J. E. G. Zetzel. Cambridge, England: Cambridge University Press.
- Crain, S., & Nakayama, M. (1987). Structure dependency in grammar formation. *Language*, 63(3), 522–543.
- Crain, S., & Pietroski, P. M. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy*, 24(2), 139–186.
- Crain, S., & Thornton, R. (2006). Acquisition of syntax and semantics. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 1073–1110). Amsterdam: Elsevier.
- Cushman, F. A., Young, L., & Hauser, M. D. (2006). The role of conscious reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science*, 17(12), 1082–1089.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford, England: Oxford University Press.
- Dunn, J. (1999). Making sense of the social world: Mindreading, emotion, and relationships. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 229–242). Mahwah, NJ: Erlbaum.
- Dupoux, E., & Jacob, P. (2007). Universal moral grammar: A critical appraisal. *Trends in Cognitive Sciences*, 11(9), 373–378.
- Dwyer, S. (1999). Moral competence. In K. Murasugi & R. Stainton (Eds.), *Philosophy and linguistics* (pp. 169–190). Boulder, CO: Westview Press.
- Dwyer, S. (2006). How good is the linguistic analogy? In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Culture and cognition* (pp. 237–256). Oxford, England: Oxford University Press.
- Dwyer, S. (2009). Moral dumbfounding and the linguistic analogy: Methodological implications for the study of moral judgment. *Mind & Language*, 24(3), 274–296.
- Dwyer, S., & Hauser, M. D. (2008). Dupoux and Jacob's moral instincts: Throwing out the baby, the bathwater and the bathtub. *Trends in Cognitive Sciences*, 12(1), 1–2.
- Eisenberg, N., Guthrie, I. K., Murphy, B. C., Shepard, S. A., Cumberland, A., & Carlo, G. (1999). Consistency and development of prosocial dispositions: A longitudinal study. *Child Development*, 70(6), 1360–1372.
- Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314.
- Feinberg, J. (1984). *The moral limits of criminal law, Vol. 1: Harms to others*. New York: Oxford University Press.
- Flanagan, O. (1991). *Varieties of moral personality*. Cambridge, MA: Harvard University Press.
- Fodor, J. (1975). *The modularity of mind*. Cambridge, MA: MIT Press.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.
- Gigerenzer, G. (2008). Moral intuition = fast and frugal heuristics? In W. Sinnott-Armstrong (Ed.), *Moral psychology. Volume 2. The cognitive science of morality: Intuition and diversity* (pp. 1–26). Cambridge, MA: MIT Press.
- Gilbert, D., & Malone, P. (1995). The correspondence bias. *Psychological Bulletin*, 117(1), 21–38.
- Greene, J. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, 11(8), 322–323.
- Greene, J., Cushman, F. A., Stewart, K. L., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111(3), 367–371.

- Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517–523.
- Greene, J., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural basis of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389–400.
- Greene, J., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2107.
- Grotius, H. (1925). *Prolegomena to The Law of War and Peace*. Translated by F. W. Kelsey et al. Oxford, England: Oxford University press.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Harman, G. (1999). Moral philosophy and linguistics. In K. Brinkman (Ed.), *Proceedings of the 20th World Congress of Philosophy* (pp. 107–115). Bowling Green, OH: Philosophy Documentation Center.
- Harris, P. L., & Núñez, M. (1996). Understanding of permission rules by preschool children. *Child Development*, 67(4), 1572–1591.
- Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Ecco Press.
- Hauser, M. D. (2009). The possibility of impossible cultures. *Nature*, 460, 190–196.
- Hauser, M. D., Chen, M. K., Chen, F., & Chuang, E. (2003). Give unto others: Genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 270, 2363–2370.
- Hauser, M. D., Cushman, F., Young, L., Jin, R. K.-X., & Mikhail, J. (2007). A dissociation between moral judgment and justification. *Mind & Language*, 22(1), 1–21.
- Hauser, M. D., & Young, L. (2008). Modules, minds and morality. In D. Pfaff, C. Kordon, P. Chanson, & Y. Christen (Eds.), *Hormones and social behavior* (pp. 1–12). Berlin: Springer Verlag.
- Hauser, M. D., Young, L., & Cushman, F. (2008). Reviving Rawls' linguistic analogy. In W. Sinnott-Armstrong (Ed.), *Moral psychology. Volume 2. The cognitive science of morality: Intuition and diversity* (pp. 107–143). Cambridge, MA: MIT Press.
- Helwig, C. C., & Turiel, E. (2002). Children's social and moral reasoning. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 476–490). Malden, MA: Blackwell.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of *Homo economicus*: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, 91(2), 73–78.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cadenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, 312(5781), 1767–1770.
- Hoffman, M. (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* (pp. 236–274). Cambridge, England: Cambridge University Press.
- Hornstein, N. (2009). *A theory of syntax: Minimal operations and universal grammar*. Cambridge, England: Cambridge University Press.
- Huebner, B., Dwyer, S., & Hauser, M. D. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13(1), 1–6.
- Hume, D. (1740/1888). *A treatise of human nature*. Ed. L. A. Selby-Bigge. Oxford, England: Clarendon Press.
- Jackendoff, R. (2007). *Language, consciousness, culture: Essays on mental structure*. Cambridge, MA: MIT Press.
- Jones, E., & Harris, V. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3(1), 1–24.
- Kamm, F. M. (2007). *Intricate ethics: Rights responsibilities and permissible harm*. Oxford, England: Oxford University Press.
- Kant, I. (1993) *Grounding of the metaphysics of morals*. Translated by J. W. Ellington. Indianapolis: Hackett.

- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63(279), 190–193.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. D., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, 446(7138), 908–911.
- Kohlberg, L. (1981/1984). *The philosophy of moral development* (Vols 1 and 2). New York: Harper & Row.
- Laurence, S., & Margolis, E. (2001). The poverty of the stimulus argument. *British Journal for the Philosophy of Science*, 52(2), 217–276.
- Leslie, A. M., Gelman, R., & Gallistel, C. R. (2008). The generative basis of natural number concepts. *Trends in Cognitive Sciences*, 12(6), 213–218.
- Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: “theory of mind” and moral judgment. *Psychological Science*, 17(5), 421–427.
- Leslie, A. M., Mallon, R., & DiCorcia, J. A. (2006). Transgressors, victims, and crybabies: Is basic moral judgment spared in autism? *Social Neuroscience*, 1(3&4), 270–283.
- Mallon, R. (2008). Reviving Rawls’s linguistic analogy inside and out. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 2: The cognitive science of morality: Intuition and diversity* (pp. 145–155). Cambridge, MA: MIT Press.
- Marcus, G. (1993). Negative evidence in language acquisition. *Cognition*, 46, 53–85.
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, 11(4), 143–152.
- Mikhail, J. (2008). Moral cognition and computational theory. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Vol. 3: The neuroscience of morality: Emotion, disease and development* (pp. 881–891). Cambridge, MA: MIT Press.
- Mikhail, J. (2009). Moral grammar and intuitive jurisprudence: A formal model of unconscious moral and legal knowledge. In D. M. Bartels, L. J. Sitka, & D. L. Martin (Eds.), *Psychology of learning and motivation, vol. 50: Moral judgment and decision making* (pp. 27–100). San Diego, CA: Academic Press.
- Mikhail, J. (2010). *Elements of moral cognition: Rawls’ linguistic analogy and the cognitive science of moral and legal Judgment*. Cambridge, England: Cambridge University press.
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional MRI investigation of basic and moral emotions. *Journal of Neuroscience*, 22(7), 2730–2736.
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2007). The neural basis of human moral cognition. *Nature Reviews: Neuroscience*, 6(10), 799–809.
- Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, 19(6), 549–557.
- Newmeyer, F. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. Oxford, England: Oxford University Press.
- Nichols, S. (2004). *Sentimental rules*. Oxford, England: Oxford University Press.
- Nucci, L. (2001). *Education in the moral domain*. Cambridge, England: Cambridge University Press.
- Núñez, M., & Harris, P. L. (1998). Psychological and deontic concepts: Separate domains or intimate connection? *Mind & Language*, 13(2), 153–170.
- Pareto, V. (1906/1971). *Manual of political economy*. Trans. A. S. Schweir & A. N. Page. New York: A. M. Kelley.
- Poeppel, D., & Embick, D. (2005). The relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 103–120). Mahwah, NJ: Erlbaum.
- Poeppel, D., & Monahan, P. J. (2008). Speech perception: Cognitive foundations and cortical implementation. *Current Directions in Psychological Science*, 17(2), 80–85.
- Prinz, J. (2006). *The emotional construction of morals*. Oxford, England: Oxford University Press.
- Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation of cognitive science*. Cambridge, MA: MIT Press.
- Quinn, W. S. (1989). Actions, intentions, and consequences: The doctrine of double effect. *Philosophy & Public Affairs*, 18(4), 334–351.

- Range, F., Horn, L., Viranyi, Z., & Huber, L. (2009). Absence of reward induced aversion to inequity in dogs. *Proceedings of the National Academy of Sciences*, 106(1), 340–345.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 337–383). New York: Academic Press.
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15(2), 165–184.
- Scholl, B. J., & Leslie, A. M. (1999). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. In E. Lepore & Z. Pylysyhn (Eds.), *What is cognitive science?* (pp. 26–73). Oxford, England: Blackwell.
- Sidgwick, H. (1907). *The methods of ethics* (7th ed.) Indianapolis: Hackett.
- Sinnott-Armstrong, W., Mallon, R., McCoy, T., & Hull, J. G. (2008). Intention, temporal order and moral judgments. *Mind & Language*, 23(1), 90–106.
- Smart, J. J. C., & Williams, B. A. O. (1973). *Utilitarianism: For and against*. Cambridge, England: Cambridge University Press.
- Smetana, J. G. (1989). Toddlers' social interactions in the context of moral and conventional transgressions in the home. *Developmental Psychology*, 25(4), 499–508.
- Smetana, J. G. (2006). Social domain theory: Consistencies and variation in children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 119–154). Mahwah, NJ: Erlbaum.
- Smetana, J. G., & Braeges, J. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly*, 36, 329–346.
- Smith, A. (1817). *The theory of moral sentiments*. Philadelphia, PA: Antony Finley.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 86–96.
- Sripada, C., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *Innateness and the structure of the mind* (Vol. 2, pp. 280–301). Oxford, England: Oxford University Press.
- Stich, S. (1972). Grammar, psychology, and indeterminacy. *Journal of Philosophy*, 69(21), 799–818.
- Sunstein, C. (2005). Moral heuristics. *Behavioral and Brain Sciences*, 28(4), 531–542.
- Thomson, J. J. (1985). The trolley problem. *Yale Law Journal*, 94(6), 1395–1415.
- Thomson, J. J. (2008). Turning the trolley. *Philosophy & Public Affairs*, 36(4), 359–374.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.
- Turiel, E. (1998). The development of morality. In N. Eisenberg (Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 863–892). New York: Wiley.
- Tversky, A., & Griffin, D. (1991). Endowment and contrast in judgments of well-being. In F. Strack, M. Argyle, & N. Schwartz (Eds.), *Subjective well being* (pp. 101–118). Oxford, England: Pergamon.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Waldman, M., & Dietrich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, 18(3), 247–253.
- Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage*, 40(4), 1912–1920.