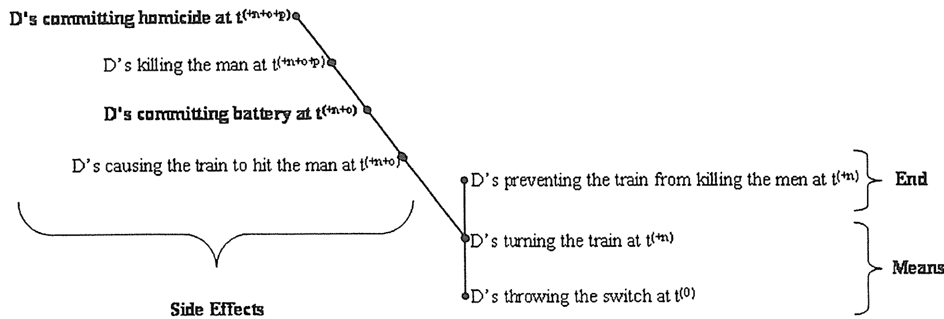


Figure 2 (Mikhail). Mental representation trolley problem (Mikhail, in press).

able; by contrast, causal reversals using "by" to connect nodes in the upward direction ("D threw the switch by turning the train," "D turned the train by killing the man") will be deemed unacceptable. Likewise, descriptions using the phrase "in order to" to connect nodes in the upward direction along the vertical chain of means and ends ("D threw the switch in order to turn the train") will be deemed acceptable. By contrast, descriptions of this type linking means with side effects ("D threw the switch in order to kill the man") will be deemed unacceptable. In short, there is an implicit geometry to these representations, which Sunstein neglects but an adequate theory can and must account for.

"The law has long used actors' intent or purpose to distinguish between two acts that may have the same result" (Vacco vs. Quill 1997, p. 802). Simple but revealing thought experiments like the footbridge and trolley problems suggest that ordinary mortals do so as well. Perhaps this explains why so many legal doctrines turn on an analysis of purpose and on the distinction between intended and foreseen effects (Mikhail 2002). Of course, some of these doctrines may constitute the kind of overgeneralization Sunstein usefully warns against. But many others presumably do not. Consider the norms of proportionality and noncombatant immunity in the law of armed conflict, which limit the permissibility of harming civilians as a side effect of an otherwise justifiable military operation and categorically prohibit directly targeting them. Are these norms the product of heuristics, or of shared principles of moral competence? The fact that we can seriously contemplate the latter alternative – that cognitive science and human rights can be linked in this manner – is significant and worth reflecting upon. In the final analysis, Sunstein's insistent homunculus may be the human sense of justice, which behaviorism in all its varieties leads us to ignore, but which we persistently disregard at our own peril.

## Do normative standards advance our understanding of moral judgment?

David A. Pizarro[a] and Eric Luis Uhlmann[b]

[a]*Department of Psychology and Social Behavior, University of California – Irvine, Irvine, CA 92697-7085;* [b]*Department of Psychology, Yale University, New Haven, CT 06520.* dpizarro@uci.edu eric.uhlmann@yale.edu

**Abstract:** Sunstein's review of research on moral heuristics is rich and informative – even without his central claim that individuals often commit moral errors. We question the value of positing such a normative moral framework for the study of moral judgment. We also propose an alternative standard for evaluating moral judgments – that of *subjective rationality*.

Sunstein wants to extend Kahneman et al.'s (1982) thesis that generally adaptive cognitive heuristics also lead to systematic and predictable errors in judgment, and makes the provocative argument that moral heuristics can "lead to mistaken and even absurd moral judgments" (target article, Abstract). Sunstein makes an important contribution to the literature on moral judgment by highlighting the role of intuitions in everyday moral thinking (see also Haidt 2001). Although Sunstein does not endorse any grand moral theory explicitly (e.g., Utilitarianism or Kantianism), he agrees that the very concept of a "moral error" requires a normative benchmark, and endorses "weak consequentialism" as being, in his view, a relatively uncontroversial standard by which to judge the successes and failures of various moral judgments.

We do not wish to debate the virtues and vices of any normative moral theory – this is a task best left to philosophers. However, we do question the necessity of positing a normative framework for understanding the psychology of moral judgment. Does a good theory of moral judgment require an objectively "right" set of moral criteria with which to compare lay judgments? Perhaps not. We believe that the research reviewed by Sunstein is extremely informative without the additional claim that individuals are making mistakes. For example, knowing and predicting the conditions under which individuals rigidly adhere to principles despite consequences is important for any successful moral theory. So the fact that individuals are willing to accept a (slightly) increased risk of dying in order to punish a betrayal is quite provocative – but does it add more value to claim that this is an error?

One possible downside of such an approach is a proliferation of error-focused work in the moral domain – a domain in which claiming an objective standard may simply lead to a whole lot of argument about which standard is right, at the expense of paying attention to the data. In our opinion, this was equally problematic with the approach of Kohlberg and his colleagues (cf. Kohlberg 1969) – a willingness to embrace a Kantian/Rawlsian theory of justice led to the questionable claim that certain individuals were at a "lower stage" of moral reasoning. Much like focusing on Kantian justice, focusing on moral errors may divert attention away from more fruitful areas of inquiry, such as (for example) cross-cultural differences in moral judgment (e.g., Haidt et al. 1993), or the emotional processes that underlie moral judgments (e.g., Pizarro 2000).

This does not mean that psychologists must abandon all talk of error in moral judgment – there is one sense of the word "error" that may still be useful in this domain. To the extent that people's moral judgments are influenced by factors that *even they perceive as irrational*, their judgments may be said to be in error (Kruglanski 1989). Empirical examples of this *subjective irrationality* in moral judgment are already available. For example, people believe that they punish to deter future criminals, yet their judgments are driven by the severity of the crime, not deterrence-related variables (Carlsmith et al. 2002; Sunstein refers to this as the "moral outrage" heuristic). Presumably, if a participant in this research was aware of this influence she would revise her judgment, as it fails to match her own standard.