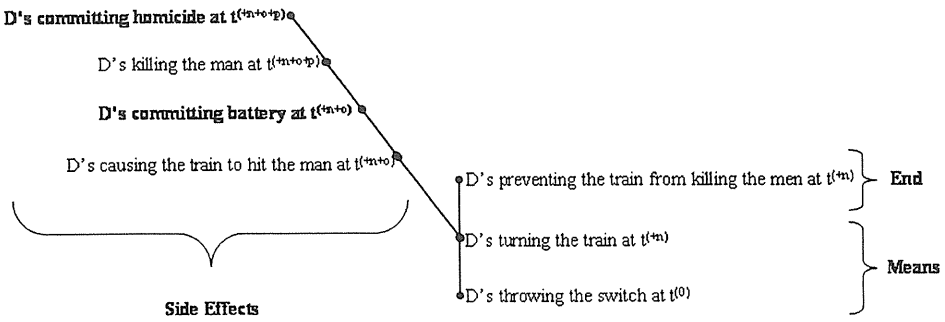

Figure 2 (Mikhail). Mental representation trolley problem (Mikhail, in press).

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.

In another study, Pizarro et al. (2003) found that participants discounted blame for intentional actions that were not carried out quite as intended (i.e., acts that lacked "intentions-in-action": Searle 1983). For example, when a murderer tripped and accidentally stabbed his victim in the process of attempting to kill him, he was perceived as less blameworthy. Interestingly, when asked to give their most rational response, participants judged acts that did and did not possess intention-in-action to be equally blameworthy. This suggests that, at least for some, discounting blame for acts that lacked intention-in-action was subjectively irrational.

In another example, Tetlock et al. (under review) examined conservative and liberal managers' reactions to a hypothetical employee error (failure to mail a package on time) with either mild or severe consequences. Both conservative and liberal managers judged the employee more harshly when the consequences of the error were severe (this has been referred to as an "outcome bias" and "moral luck"; Baron & Hershey 1988). Liberals viewed this outcome bias as an error, and reduced their recommended punishment in the severe consequences case when asked to consider how they would have reacted had the consequences been mild. In contrast, conservatives saw it as perfectly appropriate to determine the employee's punishment based on the consequences of his or her actions.

Liberals and conservatives also disagree regarding whether certain socialized intuitions are rational. Ingenious studies by Jonathan Haidt and his colleagues demonstrate that most people find it intuitively wrong to wash one's toilet with the American flag, eat one's recently expired pet, or masturbate into a dead chicken (Haidt 2001; Haidt et al. 2003). When asked to make the most rational judgment possible, liberals appear to correct for their intuitions – reducing blame for eating Fido, for example (Uhlmann et al., in preparation; see also Haidt & Hersh 2001). In contrast, conservatives provide essentially the same judgments when asked to respond rationally versus intuitively. For liberals, the judgments identified by Haidt exert a subjectively irrational influence on their judgments. But for conservatives, who place a high priority on traditional values, such judgments may seem perfectly well-grounded.

If people are indeed exhibiting "absurd moral judgments" (target article, Abstract), we suggest that this is not because heuristics lead individuals' moral judgments to diverge from some objective standard of morality (such as weak consequentialism), but because these judgments would be deemed irrational by the participant himself upon reflection. Perhaps this sense of the term "error" may be the best way to avoid the morass of subjectivity inherent in studying the moral judgments of other people, and may also keep researchers from hurling insults at each other's normative theories of choice.

ACKNOWLEDGMENT

We thank Andy Poehlman for his comments on an earlier draft of this article.

Cognitive heuristics and deontological rules

Ilana Ritov

School of Education, Hebrew University, 91905 Jerusalem, Israel.
msiritov@mscc.huji.ac.il

Abstract: Preferences for options that do not secure optimal outcomes, like the ones catalogued by Sunstein, derive from two sources: cognitive heuristics and deontological rules. Although rules may stem from automatic affective reactions, they are deliberately maintained. Because strongly held convictions have important behavioral implications, it may be useful to regard cognitive heuristics and deontological rules as separate sources of nonconsequential judgment in the moral domain.

The idea of error-prone heuristics is especially controversial in the moral domain, as Sunstein notes, although examples of choices that violate consequential principles are abundant. Among those examples are the "punishment" of companies for cost–benefit analyses to determine their investment in safety, the betrayal aversion, the resistance to "tampering with nature," and the rejection of probability of detection as a normative factor in determining punitive damages. These choices have grave consequences for the lives and well-being of many people, and the contribution of this article in drawing attention to these problems is highly important.

To ascertain that those nonconsequential judgments result from the application of mental heuristics, it is necessary to address the question of what a heuristic is. The notion of a heuristic is not well defined in the psychological literature. As Sunstein notes, Tversky and Kahneman (1984) used the term heuristic to refer to a strategy that "relies on a natural assessment to produce an estimation or a prediction." These strategies take on the form of mental shortcuts, or general purpose rules, often applied without consciousness, in judgmental tasks requiring assessment of unknown values. More recently, the evolving research on dual process theories led to a broader view of the nature of heuristics. Heuristics have come to be equated with processes of System I. This system, also referred to as the experiential system, operates automatically and effortlessly, is oriented to concrete images, and responds affectively. By contrast, the rational system, or System II, operates consciously and effortfully, and is deliberate and reason-oriented (Epstein & Pacini 1999).

In the current broad view of heuristics, not only are estimates of quantities by rules of thumb seen as the products of heuristics, but any expression of preference derived through the experiential system is regarded as such, as well. Although the boundaries of the set have not been explicitly delineated, the most notable feature of a heuristic process that distinguishes it from the cognitive processes classified as reasoning or rational is its nondeliberative nature. Although the outcomes of a heuristic can be deliberately adopted by System II, judgment by heuristic is typically an intuitive and unintentional process (Kahneman & Frederick 2002). It is usually passive and preconscious.

Returning to the examples discussed by Sunstein in the present article, these can arguably be roughly classified into two kinds: the ones that reflect the use of general cognitive heuristics in judgments (applied in the moral domain), and others that deliberate application of rules. The clearest example of a non-deliberative heuristic is the outrage heuristic in punishment. Although people are certainly aware of their outrage, they are most likely not aware of using this emotional reaction as the primary, or even the sole determinant of the punishment they set.

The resistance to cloning, stemming from the conviction that one should not "play god," or "tamper with nature," is an example of the second kind. Although the belief itself may stem from an emotional reaction, it is explicitly adopted by the rational system. The principle is held consciously and deliberately. It is relatively abstract and context-general. Similarly, the rejection of the role of probability of detection in setting punitive damages is the result of deliberate processing, often by expert and sophisticated respondents (Sunstein et al. 2000). In both of these examples, as well as in other ones, the judgment is determined by a deontological rule.

Deontological rules are rules that concern actions rather than consequences. These rules are often associated with values that people think of as absolute, not to be traded off for anything else (Baron & Spranca 1997). These protected values, compared to values that are not absolute in this way, have various predicted properties, such as insensitivity to quantity: The amount of the harm done when they are violated does not matter as much as for other values. Furthermore, in judgments involving a deontological rule or a protected value, the participation of the actor is crucial, even when the consequences are the same. The tendency to punish companies that base their decisions on cost–benefit analysis, even if a high valuation is placed on human life, may reflect the agent relativity characteristic of the rule "do not trade human life for money."