

Commentary on Sunstein: Betrayal Aversion is Reasonable

Abstract (46 words)

We accept the reality of moral heuristics in judgment and decision making, but caution that it may be hard to establish that these heuristics produce systematic errors. We argue that the morality heuristic people apply to avoid the possibility of safety product betrayal may be reasonable.

Text (758 words)

The premise that people often rely on broad, simple, moral intuitions for making judgments seems reasonable enough. Indeed, in this era where the international war on terror dominates nightly news coverage, it would be strange if moral intuitions did *not* impact the decisions people make. However, it is less clear that these moral intuitions – or moral heuristics – are as prone to systematic error as the classic heuristics (availability, representativeness, and anchoring and adjustment) described by Amos Tversky and Daniel Kahneman (Tversky & Kahneman, 1974). Tversky and Kahneman leaned on the authority of logic and probability theory – often in combination with some controversial assumptions about how people represented the tasks – to illustrate the occasional failures of their heuristics. However, the normative status of moral heuristics is less grounded. Although this shortcoming does not make moral heuristics any less real or important than other heuristics, it does mean that there is plenty of room to challenge Sunstein’s own judgments about when moral heuristics have misfired.

Consider the betrayal aversion phenomenon Sunstein discusses at V, A, 3. The research we performed on this phenomenon, and which Sunstein reviews, suggests that people are willing incur great costs to avoid betrayals and seek to punish betrayals severely when they arise (Koehler & Gershoff, 2003). We provided experimental support for these phenomena in contexts where human actors betrayed and in contexts where safety products “betrayed” by causing the very harm that they were designed to prevent. Sunstein states that the morality heuristic behind betrayal aversion is “Punish, and do not reward, betrayals of trust” (p. XX). We agree that a rule along these lines operates, though we would tinker with this wording and describe the heuristic as “Avoid and punish betrayals of trust.”

We also agree with Sunstein that the morality heuristic appears to work well in cases involving betrayals by human actors. A security guard who commits an act of betrayal by robbing the store he is paid to protect deserves the tough punishment he will no doubt receive because his crime causes multiple harms. His betrayal not only causes the focal harm to the business, but it also damages the victims’ ability to trust other security officers and undermines their sense of the social order.

However, we are not convinced that the morality heuristic “misfires” when used in response to safety product betrayals. Most participants in our study indicated that they

were willing to double their risk of dying (from 1% to 2%) to eliminate an even smaller risk of dying as a result of a betrayal (Koehler & Gershoff, 2003). For example, most people preferred an airbag that carried with it a 2% risk of dying in a serious automobile crash rather to one that carried with it a 1% risk plus an additional 0.01% risk due to fatal deployment of the airbag. Sunstein thinks this is an example where the moral heuristic “punish, and do not reward, betrayals of trust” misfires and leads to error. A safety product should be chosen, Sunstein says, “if and only if it decreases aggregate risks.”

The decision rule that Sunstein favors presumes that a safety product betrayal, unlike a betrayal by a human actor, does not produce multiple harms. We respectfully disagree. Though a safety product lacks the intentionality of a human actor, the negative consequences of a safety product betrayal may be as varied, severe, and protracted as other types of betrayals. An airbag that kills drivers who would otherwise survive the car accident can instill a deep mistrust of car manufacturers and government safety agencies among the victims’ families and friends. Safety products that betray people in this manner – by causing the very harm we trust them to prevent – may also increase our sense of vulnerability in the world and arouse a variety of negative emotions. If the negative consequences of safety product betrayals reach beyond the immediate harm, then it is not clear that people’s safety product preferences should be judged against a benchmark that only considers aggregate risks of the immediate harm. Indeed, we thought that our results were striking not because they showed how an otherwise reasonable heuristic could lead to absurd preferences, but because they indicated that the consequences of betrayal are so unbearable that people are willing to incur substantial costs to eliminate them. Certainly it would be unreasonable if people chose to avoid betrayal risks at *all* costs. But we are not persuaded that a finding that people are willing to incur *some* additional cost to avoid betrayal provides evidence of a heuristic gone awry.

References

Koehler, J. J., & Gershoff, A. D. (2003). Betrayal aversion: When agents of protection become agents of harm. Organizational Behavioral and Human Decision Processes, 90, 244-261.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1130.