Journal of Cognition and Neuroethics

Moral Heuristics and Biases

Mark Herman
Bowling Green State University

Biography
Mark H. Herman is a graduate student in philosophy at Bowling Green State University. His primary research interest is the applicability of the heuristics and biases paradigm to moral psychology. His research interests include moral psychology, ethics, rationality, and philosophy of science. He was born and raised in New York City, and received his BA from Wesleyan University. He can be contacted at hermanm_at_bgsu.edu.

Moral Heuristics and Biases

Mark Herman

Abstract

The cognitive heuristics and biases research program demonstrates the explanatory success of the heuristics-and-biases model. Expanding applications of the model to include moral judgment and action would be in keeping with the model's history of expansion and would further various scientific, philosophical, and practical interests. However, such an application faces a major philosophical challenge: justifying the ascription of "moral erroneousness" within the context of a scientific research program. In other applications of the model, staple bases for ascribing erroneousness include incongruity with empirical facts and violations of ideal theoretical rationality. However, these standards cannot be adapted to the moral domain; this is because as "objective" standards, their moral analogs would involve scientifically inadmissible appeals to "objective" morality. Yet, the previously fruitful use of some "subjective" standards suggests that, at least in principle, a "subjective" moral standard could undergird a fruitful "moral heuristic (-analogues) and biases" research program. While this "subjectivist opening" is promising, it also raises a full slate of further questions for future work.

> One death is a tragedy; a million deaths: a statistic.
> —Joseph Stalin (attributed)

Most people agree that racial discrimination is morally wrong. Despite this, there are people who discriminate on the basis of race. Why do they do this? To illuminate the project at hand, let us temporarily stipulate that racial discrimination "is morally wrong" (and briefly set aside metaethical worries). With this stipulation, we can ask, "Why do people commit this immoral act?" and ask more generally, "Why do people commit immoral acts?"

Here are some possible explanations. (1) They have the wrong moral view (i.e., they subscribe to a putative morality that permits immoral acts). (2) They do not care (enough) about morality (i.e., the weight they assign to moral considerations is insufficient to preclude decisions to act immorally). And (3) they cannot control themselves (i.e., while they do not sanction their performance of immoral acts, they succumb to impulses—that is, they are weak-willed or *akratic*).

Presumably, there is not a single explanation of all immoral acts. Instead, there is a multitude of explanations, with each explanation applying to its own particular set of cases. Another explanation (or hypothesis) that might merit inclusion amongst this multitude is one that focuses on moral reasoning and appeals to "moral heuristic (-analogues) and biases."

In the following, I will first, provide a brief primer on heuristics and biases. Second, I will discuss the expanding application of the heuristics-and-biases model and the fittingness of extending that expansion to include moral judgment and action. Third, I will identify potential benefits of that extension and provide some sense of what the application might involve. Fourth, I will introduce a major philosophical challenge confronting this application, namely, justifying the ascription of "moral erroneousness" within the context of a scientific research program. Fifth, I will attempt to make some progress on that challenge and lay out an agenda for future work.

Heuristics & Biases

The notion of "*moral* heuristic (-analogues) and biases" presupposes the conceptual framework of *cognitive* heuristics and biases. Any discussion of the former requires some familiarity with the latter. The following overview, though simplified, should be sufficient.

Humans are constantly making judgments. Some of our judgments are correct; some are erroneous. Some of these errors in judgment conform to patterns. For example, when we are asked to estimate the probability of various events, we tend to *over*estimate the probability of *dramatic* events (e.g., plane crashes) (Tversky and Kahneman 1982a). Errors that fit such patterns are *systematic*; such systematic errors constitute *biases*. The tendency to overestimate the probability of dramatic events is a type of cognitive bias— specifically, a subtype of *availability bias*.

Judgments are the product of cognitive processes. The above judgments regarding the probability of events are the product of a cognitive process that infers the probability of an event *by* assessing how easy it is to recall similar events. In other words, the process uses ease-of-recall as a proxy for probability. Dramatic events, such as plane crashes, are disproportionately easy to recall. As such, assessing ease-of-recall to infer a dramatic event's probability leads to an overestimate.

We can assess ease-of-recall more efficiently than we can assess probability (as the latter would require executing arduous algorithms). Thus, using the former to infer the latter constitutes a shortcut. Such shortcuts are *heuristics*. Since an event's ease-of-recall constitutes its *availability*, the process in the above case is called the *availability heuristic*.

Heuristics and biases that involve inferences of this sort are known as "*cognitive*" heuristics and biases.[1]

In sum, cognitive heuristics are distinct processes of unemotional inference-making (of the above sort) that constitute cognitive shortcuts (vis-à-vis executing ideally rational algorithms). Cognitive heuristics typically yield practical advantages, such as rapid judgment production; however, the judgments they produce are prone to systematic error, that is, *cognitive biases*.

Expanding Application of the Heuristics-and-Biases Model

The principal architects of the cognitive heuristics and biases research program were Daniel Kahneman and Amos Tversky (e.g., Kahneman, Slovic and Tversky 1982; Kahneman and Tversky 1996; 2000; Tversky and Kahneman 2003). The cognitive heuristics and biases model that they employed was adapted from similar models concerning *perceptual* judgments. A particularly influential case was Egon Brunswik's (1943) application of his lens model to the haze illusion (Kahneman and Frederick 2002, 52; Tversky and Kahneman 2003, 35). The haze illusion is an optical illusion in which objects seen in hazy weather appear farther away than they actually are. This phenomenon was captured as the "perceptual bias" of systematically overestimating the distance of objects in such weather. It was explained by the "perceptual heuristic" of assessing the distance of an object by assessing the perceived clarity of its edges (and the fact that in such weather, perceived edge clarity decreases). This perceptual heritage of Kahneman and Tversky's model is reflected in the occasional reference to their subject matter as "cognitive *illusions*" (Kahneman and Tversky 1996).

As Kahneman and Tversky's program progressed, applications of the cognitive heuristics and biases model expanded. This included expanding the range of phenomena captured by the model, and increasing the extent of capture (e.g., advancing from vague awareness of a tendency, to identifying a systematic bias, to hypothesizing a responsible heuristic, to empirically corroborating the heuristic's existence).

---

1. The term, "cognitive," may be somewhat of a misnomer. In psychology, "cognitive" often regards information-processing, *simpliciter*, and encompasses a variety of capacities, including visual and linguistic capacities (e.g., Anderson 2005; Balota and Marsh, 2004; Eysenck 2001; Kellogg 2002). However, in the context of the "*cognitive*" heuristics and biases program, "cognitive" connotes "cold" (i.e., unemotional) "thinking" or "reasoning," as exemplified by inferences regarding the probability of events. In this sense, the "cognitive" is distinct from other kinds of information-processing.

Just as Kahneman and Tversky adapted the "*perceptual* heuristics and biases" model and applied it to probabilistic judgments, Nisbett and Ross (1980) in turn, adapted Kahneman and Tversky's model and applied it to psychosocial judgments. For example, Nisbett and Ross identified the "act-observer bias," in which assessments of others' behavior overestimate the influence of their character (and underestimate the influence of their circumstances). Researchers working in the Kahneman and Tversky tradition went on to (adapt and) apply the heuristics-and-biases model to an expanding range of phenomena, including syllogistic inferences, causal attribution, memory, prediction, and psychophysical judgments. As the presumed bright line dividing reason from emotion was revealed to be more and more blurry, applications of the model expanded beyond solely "cold" (unemotional) reasoning. It incorporated emotion-laden processing, such as the affect heuristic, and reinterpreted "motivated" or "hot" (emotional) irrationality, such as wishful thinking. The expansion of the model's application to include emotion-induced irrational judgments supports the prospect of further expansion to include *akrasia*-induced irrational judgments and actions. Recent developments in akrasia research (e.g., Baumeister and Tierney 2011) appear to render this area increasingly ripe for the identification of biases. Another area of progress is the application of the model to assessments of value (e.g., sunk-cost bias, framing effects, and status-quo bias) (Baron 2008; Connolly, Arkes and Hammond 2000; Gilovich, Griffin and Kahneman 2002; Koehler and Harvey 2004; Schneider and Shanteau 2003).

Importantly, the range of judgments now captured under the heuristics-and-biases model include: emotion-laden judgments, value assessments, and judgments regarding the social domain. These judgments include "ingredients" that in combination, seemingly amount to something fairly close to a moral judgment. Applying a variant of the heuristics-and-biases model to moral judgments (and resultant acts) would be in keeping with the model's history of increasingly expanding its application.

Moral Heuristics (-Analogues) & Biases

A successful application of the heuristics-and-biases model to the moral domain could further various scientific, philosophical, and practical interests. For one, it would enrich our understanding of human morality—a subject matter of great import and interest. It also would unify disparate, but seemingly related phenomena currently scattered across the literature (such as displaced aggression and, as elaborated below, the identifiable victim effect). It would provide a framework with which similar phenomena could be discovered and/or incorporated into scientific study. And, it could yield implications

for moral and legal philosophy. For instance, it could identify certain immoral acts as attributable to faulty processing (as opposed to for instance, moral indifference); this would have implications regarding moral responsibility.

Perhaps most importantly, a successful program could yield practical benefits. Other applications of the heuristics-and-biases model have yielded *de-biasing* techniques for reducing error. These techniques have targeted biases in: (a) perceptual judgment (e.g., pilots' susceptibility to autokenesis illusions [Civil Aviation Authority 2002]), (b) cognitive judgment (e.g., confirmation bias amongst law enforcement officers [Rossmo 2006]), and (c) psychosocial judgment (e.g., overly-attributing others' actions to character vis-à-vis circumstances [Tetlock 1997]). Applying the heuristics-and-biases model to the moral domain could yield similar preventative and corrective measures to reduce "erroneous" moral judgments, decisions, and actions. These measures could supplement organized efforts and policies to reduce "immoral" actions. Such measures might also have applications within contemporary character education.

Even if the application of the heuristics-and-biases model to the moral domain was unsuccessful, the attempt would still be theoretically constructive. Considering all of the model's fruitful applications, the application of the model itself constitutes a progressive scientific research program (in the Lakatosian sense [Lakatos 2000]). This renders an attempted application to the moral domain a fitting test of the program's boundaries.

Applying the heuristics-and-biases model to the moral domain may reveal the following. (1) Some immoral acts can be explained as instantiations of systematic *moral* errors—that is, "moral biases." (2) Some immoral acts that instantiate moral biases are the product of and thus, explained by "moral heuristics"—that is, distinct psychological processes that constitute moral reasoning shortcuts. Alternatively, (2a) those explanatory processes might be distinguished, not by constituting a shortcut, but by possessing an alternative property that fulfills the same explanatory functions; we could call such processes "moral heuristic-analogues."[2]

Examples of phenomena that might involve "moral heuristic (-analogues) and/or biases" include: (a) implicit biases (Saul 2013), (b) the influence of displaced aggression upon prejudicial violence (Hovland and Sears 1940), (c) *psychic numbing* (i.e., indifference to the plight of individuals who are 'one of many' in a much greater problem) and its role

---

2. The explanatory functions of constituting-a-shortcut include (a) endowing heuristics with a telos (and thus, a plausible evolutionary and/or cultural origin) and (b) providing a basis for subsuming similar processes under a scientific kind.

in tolerating genocide (Slovic 2007), and (d) *blaming the victim* (Lerner and Simmons 1966).

Another such phenomenon is the *identifiable victim effect* (Small and Loewenstein 2003). This effect accounts for people's tendency to make greater charitable donations when requests are accompanied by (a) images of particular victims, versus (b) factual statements, even when such statements raise the utility of donating. This difference in donations plausibly constitutes a "moral error;" the systematicity of these decisions would render those "moral errors" instantiations of a "moral bias." A plausible explanation of this "moral bias" is a moral heuristic. Recall that the availability heuristic explained the overestimation of dramatic events by revealing that the estimates were a function not of *likelihood-of-occurring*, but the generally useful, but error-prone proxy, *ease-of-recall*. In this case, a plausible explanation of this difference in donations is that the donations are a function not of *recipients' need* (or something of that sort), but a useful, but error-prone proxy, *emotional impact of the request* (or something of that sort).

It is worth clarifying that the moral erroneousness of interest is error that is "meaningfully *moral*." While this is a slippery notion, some footing can be found by contrasting it with merely logical error committed with moral content. For instance, suppose someone makes a logical error, such as affirming-the-consequent, when responding to a moral dilemma; this would be no more meaningfully a *moral* error, than an irrational gamble on the Kentucky Derby would be an *equine* error. For an error to be meaningfully *moral*, in this sense, it must involve a substantively *moral* mistake. For instance (perhaps), one's attributing moral worth to species of animals on the basis of their cuteness.[3]

Challenge: Justifying "Moral Erroneousness"

Applying the heuristics-and-biases model to the moral domain faces challenges. One major challenge is: within the context of a *scientific* research program, on what basis

---

3. Some errors may be attributable to the use of moral content (*à la* Cosmides 1989). However, this is insufficient for being a *meaningfully moral* error. For example, suppose we have a moral reasoning module and it is the only module that affirms-the-consequent. As such, affirming-the-consequent would have an important and interesting association with moral judgment. Nevertheless, such errors would be moral in a different way from those involving *substantively moral* mistakes. It may be that all errors can be interpreted in a way that renders them ultimately reducible to logical errors. Nevertheless, I think we would still be able to make meaningful distinctions between the moral meaningfulness of (a) attributing moral worth on the basis of cuteness, versus (b) affirming-the-consequent with moral content.

can we identify an act or judgment as "morally *erroneous*?" To illuminate this challenge, consider the following.

Identifying a judgment as erroneous implies a standard from which the judgment deviates. With respect to cognitive biases, that standard is ideal rationality. For example, regarding the availability bias, identifying probability estimates as *over*estimates implies a standard of accurate probability estimation; that accuracy is per the dictates of ideal rationality (specifically, probability theory). The essence of the erroneousness ascribed by the availability *bias* is illustrated in the following figure (by the deviation of human judgment from ideally rational judgment).

Availability Bias (simplification)
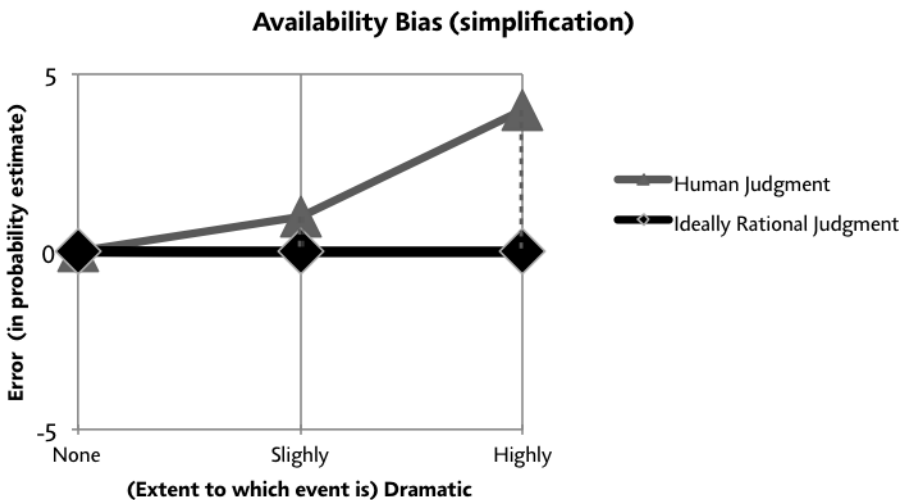
In contrast with this exemplar, from what standard might a judgment's deviation warrant the ascription, "*moral* error," within a *scientific* research program?

To aid the search for an adequate moral standard, it will be useful to consider unproblematic standards from other applications of the heuristics-and-biases model. One unproblematic standard is consistency with empirical facts. Examples of such facts include: the distance of an object (as appealed to by the haze illusion), and subsequent states-of-affairs (as appealed to by various prediction biases, such as the planning fallacy– Buehler, Griffin and Ross 2002).

The unproblematic nature of these error-ascriptions cannot be realized in a scientific research program concerned with moral erroneousness. The above error-ascriptions are for

judgments' inconsistency with facts that ultimately, are inferred due to their providing the most *scientifically epistemically* virtuous account of observations. There is no observation for which a supposed *moral* fact provides the most scientifically epistemically virtuous account to infer *within the context of a scientific* research program.[4]

Another unproblematic standard is deviation from ideal theoretical rationality. For example, error is ascribed to violations of logical principles, such as the conjunction rule (Tversky and Kahneman 1982b). This unproblematic nature cannot be realized with ascriptions of moral erroneousness. This is because within the context of a scientific research program, the case for metaethical uncertainty and skepticism is simply too strong for an unproblematic appeal to an *a priori* morality. In other words, Kant and others' attempts to derive morality from rationality failed (or at least did not succeed to an extent that would license granting moral principles equal standing in science with principles of ideal theoretical rationality) (Kant 1997). For instance, the notion that the scientific admissibility granted to the conjunction rule be bestowed upon the categorical imperative is simply a non-starter.

One might respond that a skeptical case might also be made against ideal theoretical rationality. Perhaps it can, but many of the principles of ideal theoretical rationality that have undergirded error-ascriptions are also indispensable presuppositions of psychological research, if not all of science (e.g., *modus tollens* and the relevance of sample size). As such, rejecting more than a small minority of these principles would amount to a "self-excluding argument." By this I mean that such rejections would exclude one from the community for whom the topic at hand could be of concern (*à la* defending an interpretation of international law by appealing to solipsism). Unlike skepticism of ideal theoretical rationality, moral skepticism leaves science intact (e.g., science is compatible with moral anti-realism, moral nihilism, etc.).

In sum, in the context of a scientific research program, "moral erroneousness" cannot realize the unproblematic nature of the aforementioned error-ascriptions (i.e., per empirical facts or ideal theoretical rationality). Nevertheless, opportunities arise upon realizing that the unproblematic nature of those error-ascriptions is underwritten by their standards' being in an important sense, "objective" or "mind-independent." This sort of error-ascription is not feasible with moral erroneousness because "objective morality" is simply a non-starter within a scientific research program. However, there are unproblematic error-ascriptions in fruitful scientific research programs that rely, not upon

---

4. This does not necessarily preclude the existence of moral facts; but it does preclude appealing to supposed empirical moral facts within a scientific context.

deviating from "objective" standards, but upon deviating from ideal *practical* rationality or means-to-subjective-ends rationality. Some well-established examples include ascribing error to irrational gambles (Tversky 1969), framing-sensitivity (Baron 2008, 267–71), and scope-neglect (Kahneman and Frederick 2002, 74–75).

For instance, the erroneousness ascribed to an irrational gamble depends upon the presumption that the goal or *end* of such gambling is maximizing expected financial payoff (or something of that sort). However, what grounds privileging this end over ends such as *minimizing* financial payoff, befuddling descriptive rational choice theorists, or having fun? At least within the context of a scientific research program, the privileging of such ends cannot be grounded in the aforementioned sorts of "objective" standards. Such privileging can only be grounded by the gambler's desire to maximize his financial payoff (or something of that sort). That is, such ends are in an important sense, "subjective." Upon stipulating such an end, ascribing erroneousness to judgments, decisions, and/or actions that frustrate the realization of that end may very well be based in part, upon deviating from "objective" standards, such as ideal theoretic rationality. For example, the erroneousness of irrational gambles often depends upon principles of probability theory. Nonetheless, the overall standard that grounds the erroneousness of irrational gambles still depends upon the privileging of "subjective" ends and as such, constitutes a "subjective" standard.

One might respond that the "subjectivity" of such standards would render them scientifically inadequate. This is not the case. Roughly speaking, heuristics are distinguished from pertinent contrast class members, such as algorithms, by the property, constituting-a-shortcut. 'Shortcut' is an inherently relative concept. In the cognitive program, that which heuristics are shortcuts relative to is: the algorithms dictated by ideal rationality. A fuller expression of heuristics' distinguishing property is: shortcut-vis-à-vis-ideal-rationality. Ideal rationality does not have a causally efficacious manifestation in the intuitive judgment system; shortcut-vis-à-vis-ideal-rationality is not a causally-relevant property. With respect to the causal workings of the intuitive judgment system, heuristics are indistinguishable from other intuitive processes that happen to conform to ideal rationality (e.g., simple logical rules, such as *&-elimination*). In short, heuristics are not a natural kind (or some such appropriate analogue). That is, the distinction of heuristics and biases as kinds does not stem from the causal workings of the intuitive judgment system. Instead, it depends upon their relation to ideal rationality, and stems from our valuing ideal rationality as a standard and our choosing to incorporate this standard into a particular classification scheme. While such incorporations merit caution because of certain theoretical and methodological implications, they are ultimately, unproblematic.

They merely reflect the fact that we want to understand things we care about, and that these concerns dictate categories that do not necessarily map onto the strict metaphysical categories reflected in causation. For example, that the category, money, does not map onto such categories does not mean that we should no longer study economics (Fodor 1974). The success of the cognitive heuristics-and-biases program suggests that employing value-driven categories would not impede applying the heuristics-and-biases model to the moral domain.

All in all, the fruitful utilization of "subjective" standards in other applications of the heuristics-and-biases model supports the prospect that "subjective" standards could be similarly fruitful in applications of the model to moral judgment. In other words, we have reason to think that, at least in principle, a subjective moral standard could undergird a fruitful moral heuristic (-analogues) and biases research program.

So, given this "subjectivist opening," what should that standard be? Great care needs to be taken. Following Kohlberg's problematic application of Piaget's model to the moral domain, moral psychology has already witnessed at least one case of a research program mishandling the privileging of particular moral norms (Kohlberg 1958; Piaget 1970; per: Gilligan 1977).

Here are some questions that will need to be answered. Should the standard be based upon cultural relativism (psychology is already replete with it)?[5] But morality is notoriously controversial; is there sufficient agreement to undergird a particular cultural standard? Even if there is, would such a standard be able to distinguish between moral views that are "erroneousness" from those that are merely atypical (which opposition to slavery once was)? An inability to make such a distinction would make "debiasing" disturbingly Orwellian. Does this make such standards inherently political? Is that problematic?

Relativize to the individual instead? Would moving to the individual level allow for the generalizability that is often essential to the value of research? What would an individualist standard even consist of? Moral principles? Moral competences (*à la* Mikhail's incorporation of competence vis-à-vis performance–Chomsky 1980; Mikhail 2005)? There is reason to doubt whether individuals' moral principles (rules or competences) are stable (Zimbardo 2007), genuine (Lichtenstein and Slovic 2006), relevant (Haidt 2001),

---

5. For example, the discipline's most authoritative definition of mental disorder (i.e., per the DSM-IV), includes the caveat that behaviors in support of a diagnosis "must not be merely an expectable and *culturally sanctioned* response to a particular event, for example, the death of a loved one" (emphasis added, American Psychiatric Association 2000, xxxi).

accessible (Fischhoff 1991), or even present in any coherent form (Churchland 1996). Should we appeal to the result of some procedure, such as reflective equilibrium (*à la* Rawls 1975)?

Should we instead turn our focus to privileged counterfactual judgments, decisions, and/or actions? If so, which ones? Those of particular deliberative processes (*à la* Singer 2005)? Particular intuitive processes (*à la* Kass 1998)? Those of an idealized self (*à la* Railton 1984)? If so, what kind of idealized self (e.g., Rosati 1996), and under what conditions (e.g., Firth 1952)? Furthermore, how do we operationalize such counterfactual judgments?

In addition, do instances of self-aware immorality (i.e., evil) constitute "moral *errors*"? Should "moral laziness" be excluded? Should errors-of-omission be excluded? Should the ascription of moral error be limited to explicitly *moral* judgments or include non-moral judgments that deviate from "moral correctness"? And so on.[6]

As one can see, identifying and justifying a standard for ascribing "moral erroneousness" that would be adequate in a scientific research program is a substantial challenge. While the "subjectivist opening" allows for some progress in surmounting this hurdle, there is still much work to be done.

All in all, extending applications of the heuristics-and-biases model to include moral judgments and acts would be in keeping with precedent and offers various benefits. Nevertheless, such an application faces serious philosophical challenges that necessitate future work.

---

6. The approach I favor is privileging counter-factual judgments. One possible (and preliminary) standard is that judgments are "moral errors" if they meet the following conditions: (1) they are produced and maintained *because of* the agent's lack of awareness of morally pertinent causes of the judgment, and (2) counterfactually, had the agent had this awareness, it would have prevented or overridden the judgment *because of* the moral beliefs or values of the agent. One benefit of this standard is that preventing these "moral errors" would not require any change in actors' moral beliefs, values, or motivation.

References

American Psychiatric Association. 2000. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revision). Washington D.C.: American Psychiatric Association.

Baron, Jonathan. 2008. *Thinking and Deciding*, 4th ed. New York: Cambridge University Press.

Baumeister, Roy F., and John Tierney. 2011. *Willpower*. New York: Penguin.

Brunswik, Egon. 1943. "Organismic Achievement and Environmental Probability." *Psychological Review* 50: 255–72.

Buehler, Roger, Dale Griffin, and Michael Ross. 2002. "Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman, 250–70. New York: Cambridge University Press.

Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.

Churchland, Paul M. 1996. "The Neural Representation of the Social World." In *Mind and Morals*, edited by Larry May, Marilyn Friedman, and Andy Clark, 91–108. Cambridge: MIT Press.

Civil Aviation Authority. 2002. Fundamental Human Factors Concepts. CAP 719. Safety Regulation Group. http://www.caa.co.uk/docs/33/CAP719.PDF

Connolly, Terry, Hal R. Arkes, and Kenneth R. Hammond, eds. 2000. *Judgment and Decision Making: An Interdisciplinary Reader* (2nd Ed.). New York: Cambridge University Press.

Cosmides, Linda. 1989. "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task." *Cognition* 31: 187–276.

Firth, Roderick. 1952. "Ethical Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research* 12: 317–45.

Fischhoff, Baruch. 1991. "Value Elicitation: Is there Anything in There?" *American Psychologist* 46: 835–47.

Fodor, Jerry A. 1974. "Special Sciences: The Disunity of Science as a Working Hypothesis." *Synthese* 28: 97–115.

Gigerenzer, Gerd. 1996. "On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky." *Psychological Review* 103: 592–96.

Gigerenzer, Gerd. 1998. "Surrogates for Theories." *Theory & Psychology* 8: 195–204.

Gilligan, Carol. 1977. "In a Different Voice: Women's Conceptions of the Self and of Morality." *Harvard Educational Review* 47: 481–517.

Gilovich, Thomas, and Dale Griffin. 2002. "Heuristics and Biases: Then and Now." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman, 1–17. New York: Cambridge University Press.

Haidt, Jonathan. 2001. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review,* 108: 814–34.

Hovland, Carl I., and Robert R. Sears. 1940. "Minor Studies of Aggression: Correlation of Lynchings with Economic Indices." *Journal of Psychology* 9: 301–10.

Kahneman, Daniel, and Shane Frederick. 2002. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman, 49–81. New York: Cambridge University Press.

Kahneman, Daniel, Paul Slovic, and Amos Tversky, eds. 1982. *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky. 1996. "On the Reality of Cognitive Illusions." *Psychological Review* 103: 582–91.

Kahneman, Daniel, and Amos Tversky, eds. 2000. *Choices, Values, and Frames.* New York: Cambridge University Press.

Kant, Immanuel. 1997. *Groundwork of the Metaphysics of Morals*. Mary Gregor, trans. New York: Cambridge University Press. Originally published 1785.

Kass, Leon R. 1998. "The Wisdom of Repugnance." In *The Ethics of Human Cloning*, edited by Leon Kass and John Q. Wilson. Washington, DC: American Enterprise Institute.

Koehler, Derek. J., and Nigel Harvey, eds. 2004. *Blackwell Handbook of Judgment and Decision Making*. Malden: Blackwell.

Kohlberg, Lawrence. 1958. *The Development of Modes of Moral Thinking and Choice in the Years Ten to Sixteen.* Unpublished doctoral dissertation. University of Chicago.

Lakatos, Imre. 2000. "Falsification and the Methodology of Scientific Research Programs." In *Readings in the Philosophy of Science: From Positivism to Postmodernism*, edited by Thomas Schick, Jr., 20–25. Mountain View: Mayfield.

Lerner, Melvin, and Carolyn Simmons. 1966. "Observer's Reaction to the 'Innocent Victim': Compassion or Rejection?" *Journal of Personality and Social Psychology* 4: 203–10.

Lichtenstein, Sarah, and Paul Slovic. 2006. *The Construction of Preference*. New York: Cambridge University Press.

Mikhail, John. 2005. "Moral Heuristics or Moral Competence?" *Behavioral and Brain Sciences* 28: 557–58.

Nisbett, Richard E., and Lee Ross. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs: Prentice-Hall.

Piaget, Jean. 1970. *Genetic Epistemology*. New York: Norton.

Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13: 134–71.

Rawls, John. 1971. *A Theory of Justice.* Cambridge: Harvard University Press.

Rosati, Conni. 1996. "Internalism and the Good for a Person." *Ethics* 106: 297–326.

Rossmo, D. Kim. 2006. "Strategies to help avoid investigative failures." *FBI Law Enforcement Bulletin* 75.

Saul, Jenifer. (2013). "Jennifer Saul on Implicit Bias." *Philosophy Bites*, podcast audio, December 7, 2013, http://philosophybites.com/2013/12/jennifer-saul-on-implicit-bias.html.

Schneider, Sandra L., and James Shanteau, eds. 2003. *Emerging Perspectives on Judgment and Decision Research*. New York: Cambridge University Press.

Singer, Peter. 2005. "Ethics and Intuitions." *The Journal of Ethics* 9: 331–52.

Slovic, Paul. 2007. "'If I look at the mass I will never act': Psychic numbing and genocide." Judgment and Decision Making 2: 79–95.

Small, Deborah A., and George Loewenstein. 2003. "Helping a Victim or Helping the Victim: Altruism and Identifiability." *Journal of Risk and Uncertainty* 26: 5–16.

Tetlock, Philip E. 1997. "An Alternative Metaphor in the Study of Judgment and Choice: People as Politicians." In *Research on Judgment and Decision Making: Currents, Connections, and Controversies*, edited by William M. Goldstein & Robin M. Hogarth, 657–80. New York: Cambridge University Press.

Tversky, Amos. 1969. "Intransitivity of Preferences." *Psychological Review* 76: 31–48.

Tversky, Amos, and Daniel Kahneman. 1982a. "Availability: A Heuristic for Judging Frequency and Probability." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky, 163–178. New York: Cambridge University Press.

Tversky, Amos, and Daniel Kahneman. 1982b. "Judgments of and by Representativeness." In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky, 84–99. New York: Cambridge University Press.

Tversky, Amos, and Daniel Kahneman. 1982c. "The belief in the 'law of small numbers.'" In *Judgment under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky, 23–31. New York: Cambridge University Press.

Tversky, Amos, and Daniel Kahneman. 2003. Judgment under Uncertainty: Heuristics and Biases. In *Judgment and Decision making: An Interdisciplinary Reader*, 2nd ed., edited by Terry Connolly, Hal R. Arkes, and Kenneth R. Hammond, 35–52. New York: Cambridge University Press.

Zimbardo, Philip. 2007. *The Lucifer Effect: Understanding How Good People Turn Bad*. New York: Random House.