Author Query

AQ1 Resher 2009 is not cited. Cite, delete or move to further reading?

Bias

Daniel Moseley

The term "bias" is commonly used to refer to a strong inclination either in favor or against something. The term may be used in a pejorative sense or it may be used to describe a disposition. In the pejorative sense, the inclination that is described is construed as unfair or otherwise unjustified: "The search committee member is biased against women applicants." However, there are contexts in which "bias" has a nonpejorative use: "Lucia has a bias toward neo-expressionist paintings." The focus of this entry is on a negative sense of bias. For the purposes of this entry "bias" refers to systematic errors (Kahneman 2011: 2–3). This usage is more common in mathematics and statistics than we find in ordinary language.

The ordinary pejorative usage of "bias" is not the focus of this entry; the topic of wbiee196 RACISM is the central focus of discussion in its own entry in this volume (see the wbiee896 entry on IMPLICIT BIAS for a discussion of unconscious and automatic features of prejudiced judgment and behavior). Implicit biases have been examined with the implicit attitudes test (IAT) and have been the subject of extensive empirical research.

Varieties of Error and Bias

Errors are mistakes. We might make unique individual errors or exemplify systematic ones (biases). The latter are often more serious than individual errors. Of course, there are circumstances in which an individual error may be worse than a bias. If the President of the United States declares war in error (perhaps being mistaken about whether certain public statements would constitute the declaration of war), then that would be worse than many common biases.

The magnitude of an error depends upon its gravity and extent (Rescher 2009: 4). The gravity of an error reflects the seriousness of its implications. To err in a crossword puzzle is less grave than to err in accusing someone of a crime. The extent of an error is a matter of how far off the mark an error is. Detective Rogers may be on the hunt for the dangerous criminal Smith, who is now in Chicago. If Rogers believes that Smith is in New York, the extent of the error is less than if Rogers believes that he is in Hong Kong. However, both errors are equally grave, because Smith is still on the loose. Intuitively, the greater the magnitude of an error, the greater one's culpability for making it. However, there is disagreement about the appropriate conditions for holding one culpable for errors – an important matter for the law of wbiee576 torts (*see* TORTS).

International Encyclopedia of Ethics. Edited by Hugh LaFollette.

DOI: 10.1002/9781444367072.wbiee861

Two Kinds of Errors

Here I focus on two main forms of error: errors in thought and errors in action. Errors in thought can be subdivided into cognitive, perceptual, or affective errors. Erring in a mental calculation of an arithmetic problem is a cognitive error. Visual illusions, such as the Müller-Lyer optical illusion, provide examples of perceptual errors. Having an overly gloomy or rosy view of life is an example of an affective error. Errors in action are mistakes in achieving the ends or objectives of one's actions. These are not errors in thinking about one's ends or errors in thinking about the means of achieving one's ends. Those are errors in thought. Hereafter, "practical errors" refers to errors in action. If one makes the correct mental calculation of an arithmetic problem, but errs in writing the correct answer to the problem (perhaps after the mental arithmetic is completed someone nearby begins to shout other numbers), then a practical error has occurred.

One central topic of debate in ethics is whether moral errors are practical errors; that is, whether violating a moral principle is a violation of the objectives or the ends of one's actions. Moral rationalists (e.g., Socrates, Plato, Bishop Butler, and Kant) hold that moral principles are rational requirements; according to this view, to violate a moral principle, and thereby commit a moral error, is to perform an action that is not in one's enlightened self-interest. An important topic in ethics is whether, and to what extent, considerations of self-interest are relevant to moral judgments and moral principles. For discussion of this topic, see the following entries: AMORALIST, MORAL AGENCY, and RATIONALISM IN ETHICS.

wbiee713
wbiee294
wbiee702

Another contested issue in ethics is whether all practical errors are cognitive errors. This topic has its historical roots in Western philosophy in Socrates' defense of the view that all action is motivated by cognitive states (knowledge, judgment, or belief). Socratic intellectualism is commonly associated with his defense of the unity of the virtues (UT) thesis. UT holds that (i) the possession of one virtue entails the possession of all of them, and (ii) virtue is either identical with, or necessary and sufficient for, knowledge. UT is not just a historical relic – see Wolf (2007) for a recent defense of UT.

Contemporary ethics, for the most part, has given less attention to cognitive errors than to practical errors (although Adler 2002 argues that in addressing the question "What should I believe?" the answer will appeal to standards of evidence that are intrinsic to belief). Perhaps less attention has been given to cognitive errors in ethics because it is obvious to many that in the domain of belief it is worse to form a mistaken belief than it is to suspend one's judgment and fail to believe something true. The ancient skeptics held the latter to be a virtue, but not the former. They embraced the suspension of belief as the rationally appropriate attitude to take to everything – it is the best strategy that one can take to avoid making cognitive or practical errors. However, many people who are ignorant about important things are not aware of their ignorance. More commonly, ignorance is not a cultivated state of mind as it was among the ancient skeptics, and it is a major source of error and moral wrongdoing.

Five types of ignorance have been identified that commonly lead to moral and prudential error (LaFollette 2016: 9–11). This list highlights important kinds of ignorance and is not intended to be exhaustive. First, there is ignorance of the nature of the context in which an agent is acting: a person may believe that he is attacking someone in self-defense but, in fact, the individual who seemed threatening poses no threat at all. Second, there is ignorance of relevant history: after a decade of marriage, one might discover that one's spouse is actually a long-lost biological sibling. Third, there is ignorance of human psychology and sociology: if someone embraces a crude moral psychology, such as the view that everyone is only motivated by self-interest, then one may forego many deeply meaningful and wonderful intimate relationships with others. Fourth, there is ignorance of the nature of sociopolitical institutions, especially the ones where one resides: if one does not understand the ways that our preferences and beliefs are shaped by socioeconomic class, religious institutions, and legal systems, then one cannot control or counter their pernicious influences. Fifth, there is ignorance of basic statistical, scientific, and logical principles: if one does not understand the dangers of radiation, then one would easily be duped into buying a home next to a nuclear power plant.

In short, cognitive errors are significant to ethics and in navigating one's life. They should be avoided, and the reduction of cognitive biases that exemplify these errors are an important domain of study. There has been a surge of empirical research on biases in recent decades. In the following section, I present a framework for explaining biases.

A Framework for Explaining Biases

It is helpful to distinguish four ways of explaining biases: personal, subpersonal, situational, and systemic (Cassam 2017). The distinction between personal and subpersonal is the most philosophically controversial aspect of this framework. It is first articulated in Dennett (1969) (see Drayson 2012 for some challenges in the deployment of the distinction in philosophy of mind and psychology). In this context, the distinction should be clear enough to serve as a starting point for further inquiry. The levels of explanation, to be described in what follows, proceed by starting with a focus on traits and states of persons and then zooming out toward a more general level of social description.

Personal explanations explain systematic errors by appealing to personal qualities of individuals: carelessness, gullibility, closed-mindedness, dogmatism, and wishful thinking. One prominent form of these explanations attributes vices or bad character to the agent (Cassam 2016; Zagzebski 1996). Other types of person-level explanations maintain that the systematic error arises from the agent's beliefs, desires, or other propositional attitudes – for instance, if the bias is due to a logical fallacy.

Subpersonal explanations usually explain systematic errors by appealing to the automatic, involuntary, and unconscious operation of neurophysiological or computational mechanisms. These explanations are mechanistic and causal in a way the personal explanations are not, and the causal mechanisms that are posited are universal and not specific to the individuals that realize them (Cassam 2017). These types of explanations – for example, genetic or neurological explanations – typically

centrally feature psychological properties that are not accessible to consciousness. For example, a description of activity in the occipital lobe of the brain may be centrally featured in explaining systematic errors due to visual illusions. It would not be appropriate to praise or blame someone for exemplifying this type of neurological activity (*see* PRAISE; BLAME). wbiee863

Situational explanations explain systematic errors by appealing to contingent situational factors, such as time pressure, overwork, or fatigue (Cassam 2017). Rather than focus on features of individuals, these explanations focus on the complex and demanding situations in which individuals find themselves. These types of explanation are often found in social psychology and behavioral economics (*see* SITUATIONISM). Since some people are more vulnerable to situational factors than others, these often provide incomplete explanations. wbiee171 wbiee840

Systemic explanations explain systematic errors by appealing to features of the organizations and institutions that individuals inhabit: poverty, schools, cultural norms, laws, and policies. Systemic explanations of bias usually focus on features of organizations and institutions that lead to systematic errors. For instance, a bad education may be a systemic explanation for why a person has certain biases. Legal studies, political science, economics, anthropology, and other fields of study offer descriptions of policies and norms at the systemic level. The elimination or reduction of the systemic sources of bias would involve, at least in part, improving the condition of the least well-off in society, improving schools, improving public attitudes and norms, and improving laws. Articulating what these improvements will be and developing strategies for implementing those goals are immensely challenging tasks. Some who are skeptical of our ability to make these improvements and are skeptical about the ability of human nature to overcome our biases have suggested that we deploy biomedical technologies, such as genetic therapies, pharmaceuticals, and brain–computer interfaces, to improve human nature and overcome our biases. See the entry ENHANCEMENT, BIOMEDICAL for the possibilities and perils of technological solutions to the problems of biases. wbiee560

Situational and systemic levels of explanation may be difficult to distinguish in practice. For example, in recent years, healthcare systems in the United States have made efforts to improve the quality of healthcare and the decisions of healthcare providers by "debiasing" contexts where providers are prone to error (Cassam 2017; Fargen and Friedman 2014). According to this line of thought, it is better to avoid errors than it is to find someone to blame for them. Put differently, medical institutions have sought to create a "culture of safety" instead of a "culture of blame." This shift exemplifies an important debate about whether societies should emphasize personal responsibility to reduce cognitive errors or whether they should focus on improving situational and systemic factors.

The Attribution of Biases

People often attribute biases to others as a way of discrediting them. This practice may be pernicious if prejudice leads people to inappropriately discredit another's testimony (Ballentyne 2015; Fricker 2007). This tendency should make us cautious

in attributing bias. We should examine the reasons we have for attributing bias, including the possibility that we are being biased.

If S knows that S is biased, can that person control or correct for it? There is no simple answer here: it depends on the bias and the person. Someone might be able to use some of her own biases to correct other biases: I might be biased toward a gloomy view of life but also biased to share my friends' attitudes. If so, then making friends with people who are less gloomy may make me less gloomy. Also, if S is aware that S has the availability bias (which is a cognitive bias that involves a tendency to rely on recent or emotionally charged examples when inquiring into a topic), then S might intentionally correct for it by taking extra time when writing a paper to make sure that due diligence has been exercised in researching and care taken that not just the most recent books and articles that S has read are being cited and discussed. Other biases may be exceptionally stealthy and evade self-correction (Cassam 2016). Empirical evidence suggests that overconfidence bias may be stealthy in this manner – that is, the overconfidence bias may have features that make self-correction challenging (or perhaps even impossible). If someone has the overconfidence bias, then it may prevent that person from taking steps to be more modest and careful in their inquiries, including the inquiries into that person's own biases. Self-knowledge and engaging in reflective deliberation may be sufficient to correct some biases, but situational and systemic factors may be necessary to address the ones that we cannot see or address in ourselves. (*See* NUDGING for further discussion of the ethical considerations involved in nudging people in ways that correct for biases.)

wbiee847

Acknowledgments

I would like to thank Luc Bovens, Hugh LaFollette, Loren Lomasky, Joy Meyer, and Ram Neta for helpful feedback on earlier drafts of this entry.

wbiee713 wbiee171 wbiee560 wbiee896 wbiee294 wbiee847 wbiee863 wbiee196 wbiee702 wbiee840 wbiee576

See also: AMORALIST; BLAME; ENHANCEMENT, BIOMEDICAL; IMPLICIT BIAS; MORAL AGENCY; NUDGING; PRAISE; RACISM; RATIONALISM IN ETHICS; SITUATIONISM; TORTS

REFERENCES

Adler, Jonathan 2002. *Belief's Own Ethics*. Cambridge, MA: MIT Press.

Ballentyne, Nathan 2015. "Debunking Biased Thinkers (Including Ourselves)," *Journal of the American Philosophical Association*, vol. 1, no. 1, pp. 141–62.

Cassam, Quassim 2016. "Epistemic Vices," *Monist*, vol. 99.

Cassam, Quassim 2017. "Diagnostic Error, Overconfidence, and Self-Knowledge," *Palgrave Communications*, vol. 3.

Dennett, Daniel 1969. *Content and Consciousness*. New York: Routledge & Kegan Paul.

Drayson, Zoe 2012. "The Uses and Abuses of the Personal/Subpersonal Distinction," *Philosophical Perspectives*, vol. 26, no. 1, pp. 1–18.

Fargen, Kyle, and William Friedman 2014. "The Science of Medical Decision Making: Neurosurgery, Errors, and Personal Cognitive Strategies for Improving Quality of Care," *World Neurosurgery*, vol. 82, no. 1–21, pp. E21–E29.

Fricker, Miranda 2007. *Epistemic Injustice*. Oxford: Oxford University Press.
Kahneman, Daniel 2011. *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.
LaFollette, Hugh 2016. "The Greatest Vice?" *Journal of Practical Ethics*, vol. 4, no. 2, pp. 1–24.
Resher, Nicholas 2009. *Error (On Our Predicament When Things Go Wrong)*. Pittsburgh, PA: University of Pittsburgh Press. AQ1
Wolf, Susan 2007. "Moral Psychology and the Unity of the Virtues," *Ratio*, vol. 20, no. 2, pp. 145–67.
Zagzebski, Linda 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

FURTHER READINGS

Ariely, Daniel 2008. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York: HarperCollins.
Hansen, Katherine, Margaret Gerbasi, Alexander Todorov et al. 2014. "People Claim Objectivity After Knowingly Using Biased Strategies," *Personality and Social Psychology Bulletin*, vol. 40, no. 6, pp. 691–9.
Pronin, Emily, and Kathleen Schmidt 2012. "Claims and Denials of Bias and Their Implications for Policy," in Eldar Shafir (ed.), *The Behavioral Foundations of Public Policy*. Princeton, NJ: Princeton University Press.