Hanlon's Razor

Nathan Ballantyne and Peter H. Ditto

Penultimate version, August 2021. The final version appears in *Midwest Studies in Philosophy;* please quote only from the published version:

https://www.pdcnet.org/msp/content/msp_2021_0999_9_3_3

Abstract: "Never attribute to malice that which is adequately explained by stupidity" – so says Hanlon's Razor. This principle is designed to curb the human tendency toward explaining other people's behavior by moralizing it. We ask whether Hanlon's Razor is good or bad advice. After offering a nuanced interpretation of the principle, we critically evaluate two strategies purporting to show it is good advice. Our discussion highlights important, unsettled questions about an idea that has the potential to infuse greater humility and civility into discourse and debate.

1. Introduction

People screw up. How should we explain their errors?

Potentially helpful advice comes from a principle named Hanlon's Razor—*Never attribute to malice that which is adequately explained by stupidity*.[1] The principle is called a "razor" after Ockham's Razor, the famous rule of thumb that recommends simple explanations over complex ones, an idea designed to curb the human tendency toward metaphysical extravagance. Hanlon's Razor is meant to curb our tendency toward attributional extravagance, our proclivity to make sense of other people's bad behavior by moralizing it. The Razor has been widely noted, because it seems to encapsulate an insight into human psychology and wise judgment.

An example will demonstrate Hanlon's Razor in action. Imagine you and your neighbor disagree about a hot-button issue such as immigration policy, climate change, or the death penalty. After a conversation at a backyard barbecue, you feel strongly that she must lack basic human empathy. (How else, you wonder to yourself, could she believe those horrible things?) But then you recall the Razor's advice and realize that her wrong thinking might better be explained by misinformation she received from her favored pundits. Perhaps your neighbor is not morally defective—she could just be confused about the facts. Your reflection on the conflict, guided by the Razor, leads you to readjust your view of your neighbor. You "shave off" a moralized explanation for your neighbor's thinking.

Despite the resonance and potential usefulness of Hanlon's Razor, we know of no philosophical or scientific work that explores it directly. Our guiding question here is simple. Is Hanlon's Razor good advice or bad? The Razor falls into a genre of techniques to improve human judgment, similar to advice such as "consider the opposite," "listen to both sides," or "never judge a book by its cover." Even if the Razor is flawed in some respects, our

[1] The website Quote Investigator (2016) provides details concerning the origin and history of Hanlon's Razor.

investigation here will highlight questions about how to improve judgment using rules and principles.

Our discussion will proceed as follows. The standard formulation of the Razor leaves some crucial details unstated or unsettled and so we begin in §2 by giving a more carefully nuanced version. Then in §3 and §4 we focus on two explanations for why the Razor is good advice. The first explanation says the Razor is helpful because it improves judgmental accuracy: it helps us more accurately perceive the sources of our opponents' errors. The second explanation is that the Razor is good advice because makes us more charitable toward others: it helps us see their errors as intellectually misguided rather than morally defective, even when our generous viewpoint doesn't match reality. We argue that both explanations have serious limitations and then in §5 note how failures of self-insight can make the Razor ineffective. In §6, we conclude by returning to the thought that being accurate and charitable are both values that the principle could help people attain. A deep problem, we point out, is that the Razor can't balance these values in the way our circumstances frequently demand.

2. Sharpening Hanlon's Razor

We will ask and answer four questions about the Razor and thereby reveal the principle's essential features.

First, according to the Razor, what is it that we should be cautious about attributing to malice? While it is certainly possible to invoke malice to explain someone's correct views or good behavior, we think the Razor is normally used to explain people's errors and mistakes—namely, beliefs we perceive as mistaken or behavior we believe is irrational. Importantly, that person need not have done something that's necessarily wrong in a normative sense. It is just that their behavior deviates from what *we* think is right or reasonable—that is, we believe they disagree with us about what is true or rational. Whenever someone's error could be due to malice or stupidity, the Razor's message is to proceed cautiously, ruling out stupidity before positing malice. In other words, the Razor recommends a defeasible presumption in favor of stupidity as the explanation for others' mistakes.

Second, what is the meaning of the term "stupidity" in the Razor? Like the notion of "malice," we understand "stupidity" to cover many possibilities. Some are dispositions, including the inability to think or reason effectively, the inability to search for available evidence, various (non-moral or "cold") biases, and so on. Other kinds of "stupidity" are states or state-like: a lack of accurate information about the topic at issue, being deceived or misled, a lapse in intelligence (by an intelligent thinker), an instance of bad reasoning (by a normally good reasoner), a lapse of attention (by a normally attentive thinker), a failure to seek out relevant evidence (by a normally curious or well-informed thinker). In other words, while malice attributes another's error to a moral defect, stupidity implies instead that the cause of the error lies in an *epistemic defect*. Note that we understand epistemic defects to include circumstances that systemically induce error. In such circumstances, a person may reason flawlessly and nevertheless mess up. To avoid error and know the truth, someone requires dispositions and states for thinking well—relevant cognitive skills, good evidence, favorable circumstances,

and so on. The Razor advises us to try to explain someone's error in terms of limitations and defects within this complex of epistemic skills and states.

Third, what exactly is it to "attribute to malice" people's errors? When someone else makes what we take to be an error, the error could be explained by a moral defect. We read "malice" in the Razor broadly to include unwarranted self-interest, pride, moral corruption, lack of empathy, lack of compassion, prejudice, ill will, spite, hatred, cruelty, and the like. When we "attribute to malice" some error you make, we posit a moral defect in you. We explain your error by moralizing it—you err because you're bad.[2]

Fourth, in what kind of situations is the Razor supposed to apply? That is, when is its advice likely to be most relevant? We noted it's supposed to help when people think someone else has made a mistake, but we can say more. Sometimes attributing error to malice or stupidity is a non-starter and the Razor should not be applied. For instance, when a person fails on a straightforward task of ability, such as solving a simple arithmetic problem, observers are likely to see stupidity as an obvious and adequate explanation, with little need to consider malice as a potential explanation. And when a person fails a straightforward test of ethical judgment or moral sensitivity—such as recognizing that torturing innocent people for pleasure is wrong—observers are likely to quickly issue a malice attribution, viewing stupidity as inadequate to explain such a blatant moral failing. The Razor appears most relevant when instances of both attribution types are plausible and thus "in play." Think of the range of cases as spread out along a spectrum, where the plausibility of the relevant attributions varies from highly plausible to highly implausible. The Razor applies to cases toward the middle of the spectrum, not the extreme ends. It is generic advice but not universal in scope.

We propose to sharpen Hanlon's Razor as follows: When someone's error in belief or behavior can be plausibly explained by either malice or stupidity, *never attribute that error to a moral defect when it can be adequately explained by an epistemic one.* What this version lacks in catchiness, it makes up for in clarity. Let's now proceed to investigate our main question: Is Hanlon's Razor good or bad advice?

3. Debiasing biased attributions

The Razor may be good advice because it improves the accuracy of our judgments about others. Try a first pass on the idea. Plausibly, people are over-inclined to see the world through a moral lens (Rozin 1999; Tetlock 2003; Knobe 2010). Situations, topics, and questions that do not necessarily concern morality become "moralized." And even when issues do have legitimate moral aspects, people are too harsh in their judgments of opponents' moral character. Consequently, people often mistakenly attribute their opponents' thinking and behavior to moral defects, rather than to differences in evidence, reasoning styles, or other

[2] The term "malice" normally picks out an evil intention to harm or cause injury. On our interpretation of the Razor, the term refers to a range of morally bad features, in addition to an ill will. In support of our interpretation is the fact that people sometimes appeal to the Razor to curb attributions of prejudice, self-interest, and other states that do not require anything like an evil intention.

epistemic factors. To the extent that people's blunders are due more often to epistemic defects than moral ones, the Razor's message should make attributions more accurate. In other words, the Razor's message could be debiasing because it is de-moralizing.

We can summarize that reasoning as follows:

(1) Our negative moral attributions are systematically biased in some conflict[3] C.
(2) If our negative moral attributions are systematically biased in C, then following the Razor's guidance can make our negative moral attributions more accurate in C.
(3) Therefore, following the Razor's guidance can make our negative moral attributions more accurate in C.

Let's examine each premise. Is premise (1) plausible? Psychological research suggests that humans are inveterate moralizers—or, to use a bit of contemporary vernacular, people are *judgy*. Moral condemnation and the desire to punish moral transgressors is a cultural universal that plays a crucial role in maintaining social coordination and cooperation (Boyd and Richerson 1992; Fehr and Gachter 2002; Henrich et al. 2006). Philip Tetlock (2003), for example, characterizes people as "lay theologians," with a deep concern for upholding values they see as sacred, and whose moral sensitivities often influence judgment processes that are normatively unrelated to moral considerations, such as the evaluation of trade-offs, the use of base-rate information, and the consideration of counterfactuals (Tetlock et al. 2000). People can't easily shut off their inner-moralizer and may be prone to overattribute moral defects to others.

A number of studies reveal a tendency to attribute malice to others. Focusing on cases of moral and political disagreement, Glenn Reeder and colleagues (2005) found strong "egocentric motive attribution": people tend to attribute self-interest and other negative motives to their opponents. Negative motive attributions were "magnified among those most strongly involved in the issue," and subjects were "wary of hidden motives in the opposition and even tended to doubt that the opposition was aware of its own motives" (Reeder et al. 2005, 1508). Confidently believing our opponents are wrong leads us to view them as driven by bad motives—even when we lack good evidence showing what their motives are. In another study, Joel Walmsley and Cathal O'Madagain (2020) found that experimental subjects tend "to expect other people to be more likely to act on the worst motive attributed

[3] The term "conflict" and synonyms pick out cases where people settle upon apparently incompatible answers to a question. Take the question "Is capital punishment ever morally justified?" as an example. A conflict could involve one person believing it is, another person believing it is not, a third suspending judgment (i.e., adopting a settled stance of neither believing nor disbelieving; see Friedman 2013), and a fourth being uncertain what to think about the issue (see the discussion of "doxastic openness" in Ballantyne 2019, 111–115). Importantly, not all conflicts are genuine: sometimes conflicts are merely verbal, where a difference in expression or unrecognized miscommunication hides deep agreement concerning a question's answer (Chalmers 2011; Ballantyne 2016).

to them" (2020, 7). When subjects evaluated others' motives in the absence of specific information about which reasons other people were in fact acting upon, subjects anticipated others were moved by the worst and most unflattering motive. Even when subjects are told that others have both good and bad motives, they tend to judge the bad ones are the main motive. Walmsley and O'Madagain call this the "worst-motive fallacy."

People also tend to view their opponents' disagreeable behavior and actions as arising from intentions and free choices. For example, Joshua Knobe and collaborators have documented the "side-effect effect" (Leslie, Knobe, and Cohen 2006; Feltz 2007): when subjects learn about an action that has unintended *negative* consequences, they often see it as more intentional than an identical action that has unintended *positive* consequences. Some moral psychologists have found that when subjects blame others for an action, they often ascribe to their targets greater control and freedom (Alicke 2000; Clark et al. 2014). People make sense of opponents' actions by attributing them to their opponents' bad moral character.

Malice attributions can also be fueled by egocentric biases.[4] For instance, "naïve realism" is the tendency for people to believe their own perceptions are veridical and capture the world as it is (Ross and Ward 1996). In conflicts, people assume "I'm right, you're biased" (Cheek and Pronin forthcoming). A plausible implication is that if someone believes that his views follow from the correct moral principles or his actions flow from personal virtue, he will naturally infer that his opponents—who think and act differently—are moved by false principles and character flaws.

Biases toward unflattering attributions can be amplified by group membership and identity. Adam Waytz, Liane Young, and Jeremy Ginges (2014) found a "motive attribution asymmetry" between ingroup and outgroup members. They examined how people see the motivations of their own group compared to other groups. People tend to believe their own group is motivated more by love for their own group than by hate and also that other groups are motivated more by anger and hate than by love. As Waytz and collaborators note, "failing to recognize love as a shared motive between one's ingroup and outgroup likely exacerbates conflict" (2014, 15690). Other research has suggested that shared moral views can "become the defining aspect of social identity," meaning that morality can fuel group conflict (Böhm, Thielmann, and Hilbig 2018, 15–16). When ingroup members believe an outgroup has incorrect moral commitments or motives, outgroup members get dismissed, denigrated, and dehumanized.

It seems then there is good reason to suspect that people in disputes tend to be hyperactive malice-attributors. We note that recent experimental evidence is echoed by reports from observers of human nature in other eras. The seventeenth-century philosopher La Rochefoucauld wrote, for instance: "Our readiness to believe evil, without investigating it adequately, results from pride and laziness. We want to find the guilty party, and we do not want to go to the trouble of investigating the crime" (1678/2007, 77, V.267). People are judgmental, suspicious, cynical, and don't give opponents the benefit of the doubt. Let us grant premise (1) for now.

[4] Thanks to Jared Celniker and Victor Kumar for discussion.

How about premise (2)? If people's attributions in situations of conflict tend to be overly moralistic, leading them to attribute opponents' beliefs and behavior to malice when they are in fact a product of stupidity, then—consistent with premise (2)—the Razor's disapproval of moralistic attributions and approval of epistemic ones should in theory make judgments more accurate. But can following the Razor really reduce commonplace biases on attributions?[5]

We think it's possible the Razor's advice could push someone in the opposite direction of commonplace biases on moral attribution. A crucial empirical issue needs sorting out, though: Is that what the Razor does to someone's thinking? Yes, sometimes, we expect. Following the principle appears to be one way to rein in biased attributions, either making them less frequent or less extreme. But what is hard to know is the scope of cases where the Razor helps.

To begin to see why the issue of scope is so important, notice that the Razor can be interpreted as a hypothesis about the *actual causes* of people's thinking and behavior in conflicts. That empirical generalization must hold for the Razor to make someone's attributions more accurate across a range of conflicts. But why think stupidity is more commonplace than malice in disputes? Consider a point that chips away at the broad generalization. When two people recognize they disagree with each other, they tend to see the other side as mistaken (Cheek and Pronin forthcoming); but it is frequently implausible that *both sides* tend to err from stupidity more than malice. In fact, one side may not be mistaken at all. Thus, we have reason to doubt the generalization at issue is a correct description of why people in conflicts tend to think and behave as they do.

---

[5] We must set other questions about premise (2) to the side, but two issues deserve a brief note.

First, we are exploring the possibility that the Razor debiases people's attributions of malice. But even if your malice attributions are correct, you could still be biased in the following sense: correct malice attributions can be part of a poor explanation of someone's error. For example, your opponent could be a moral reprobate, just as you believe he is, but his repellant character is not what accounts for his mistake and, as a result, your correct malice attribution does not illuminate why he went wrong. We sidestep that issue for now.

Second, supposing the Razor can debias, by what mechanism might do that? From the armchair, the answer is unclear. Simply alerting people to the threat of a bias is typically not sufficient for them to avoid or correct for it (Wilson and Brekke 1994; Wilson, Centerbar, and Brekke 2002). The Razor must provide more than a mere warning of potential bias. To debias effectively, the principle would need to reconfigure how people form their judgments, perhaps by shifting their thinking from "fast, automatic" processes to "slow, deliberate" ones (Kahneman 2003). Consider how that might happen. Maybe the Razor helps you recognize when your malice attributions lack strong evidence. That is, thinking about the principle shows you when you haven't eliminated the possibility your opponent is stupid as opposed to bad. Alternatively, maybe thinking about the principle makes your evaluations more sensitive to the available information about your opponent. When following the principle, you pay more attention to your opponent's potential epistemic defects than you would otherwise. Finding out how the Razor debiases, if indeed it can, awaits empirical investigation.

The empirical generalization looks even more dubious after reflection on two further kinds of cases. First, people sometimes err more from badness than stupidity. In Quentin Tarantino's 2003 film *Kill Bill*, one female character remarks, "It's mercy, compassion, and forgiveness I lack—not rationality." A real-world example could involve Pol Pot, the tyrannical leader of the Khmer Rouge who perpetrated genocide in Cambodia. In such examples, malice is the correct explanation for someone's errors, meaning that trying to explain these by stupidity would be misdirected. Second, people sometimes err because of badness *and* stupidity. Some of U.S. President Richard Nixon's mistakes in the White House can be traced to both his moral corruption and his paranoia (arguably, sometimes an epistemic defect), including the botched prosecution of the "Pentagon Papers" whistleblower, Daniel Ellsberg. Nixon's directive to his White House "plumbers" to illegally break into the office of Ellsberg's psychiatrist in search of dirt on the former defense analyst was plausibly both stupid and bad.

An empirical generalization that would explain why the Razor debiases across cases—namely, that stupidity is more commonplace than malice in conflicts—is false. That's because one side in a conflict is often correct and not necessarily stupid; sometimes errors are due to badness more than stupidity; and sometimes errors are due to a mixture of badness and stupidity. So, when opponents blast each other with negative moral attributions, we can't assume those attributions tend to be inaccurate in the full range of cases where the Razor is supposed to apply. We should not presume that the Razor leads away from error in all cases.[6]

A slightly different way to put our point: even though the Razor can correct biased attributions under some logically possible conditions, that does not guarantee it improves people's judgment in most ordinary conflicts. If the Razor is good advice because it makes us less biased, we need to know more about its performance. One possibility is that the Razor is not an effective debiasing technique. Merely warning people does not eliminate or correct many types of bias and efforts to debias may even backfire (Schwartz et al. 2007). Perhaps the Razor's message can counteract powerful biases on our attributions, but how that works remains to be discovered. Even so, we can be sure that the Razor may improve accuracy in some cases while reducing accuracy in others.

We should underline an important implication of reduced accuracy in judgment: people using the Razor may be prone to exploitation. This is one "dark side" of the principle. Imagine you disagree with someone who actively misleads you about their knowledge. They present themselves as intellectually inept or uninformed; but suppose their errors are in fact explained by bad motives. Taking the Razor's guidance by trying to rule out their stupidity before attributing malice makes you too charitable—and vulnerable to abuse if you continue to interact with them. In these situations, the Razor makes you less accurate than you would be without it. But accuracy isn't everything and one possibility is that the Razor is good advice because it makes you more charitable, even at the cost of accuracy. We turn to that idea next.

[6] We suspect that the Razor's advocates do not recommend deploying the principle only in situations where stupidity is known to be more pervasive than malice. The principle is supposed to be beneficial even if its users do not know what type of situation they are in.

4. Framing conflict charitably

Could the Razor be beneficial not because it makes people more accurate but because it helps them frame conflicts more charitably? We will try a first pass on the idea. As we noted, conflicts can become moralized. This easily happens when people hold *moral convictions*—beliefs they take to be grounded in a distinction between right and wrong, in contrast to beliefs concerning non-moral claims (see Skitka et al. 2021 for a review). In general, people tend to perceive their moral convictions as universal truths, applicable across time and place (Skitka et al. 2021, 352), and conflict over moral convictions predicts greater intolerance toward opponents. For example, people prefer greater social and physical distance from those who reject their moral convictions (Skitka et al. 2005; Zaal et al. 2017), an observation made across cultures (Skitka et al. 2013). Unsurprisingly, viewing conflict through a moral lens can make partisans judge each other harshly.

Moralization can also lead people to frame conflicts in unproductive ways. Suppose two groups disagree over an economic or policy issue. They could interpret the conflict as follows: the two groups hold the same or highly similar goals and simply happen to think differently about the best means to reach those goals. But let's imagine these groups instead view their conflict as arising from significant differences in their goals and intentions. Both sides believe the other side wants something bad, or at least wants to block a good outcome. Seeing conflict as caused by divergent goals fuels negative moral attributions as well as behaviors that are not cooperative, civil, or tolerant. The Razor might help here. It tells partisans that their conflict is not necessarily due to a clash of goals but could be caused by the other side's epistemic defects. The Razor encourages a kind of "Gestalt shift," allowing people to see their opponents more charitably—even if not more accurately. The shift presumably helps people get along better.

The basic reasoning can be expressed as follows:

(1) People who see conflict as due to differences in goals and motives tend to be less cooperative, civil, and tolerant than those who do not (all other things being equal).
(2) Following the Razor's guidance can prevent you from seeing conflict as due to differences in goals and motives.
(3) Therefore, following the Razor's guidance can make you tend to be more cooperative, civil, and tolerant than those who do not (all other things being equal).

(1)–(3), though not deductively valid, purports to explain why the Razor is good advice: the principle makes us get along better with our opponents. The rule of thumb is good advice because it helps produce a good outcome.[7]

[7] On the assumption that getting along better is not valuable all by itself, notice that cooperativeness, civility, and tolerance can be instrumentally valuable for securing all sorts of social and political benefits, including democracy.

How can we evaluate the reasoning? For starters, premise (1) appears to be supported by research on moralized conflict (Skitka et al. 2021), so we will grant it here. Premise (2) appears to be sensible at least on its face. To begin to see why, notice the Razor reminds people that disagreement can be explained by epistemic defects, not just moral ones—and also that they need to rule out the former explanations before affirming the latter. Plausibly, negative *epistemic* attributions tend to be more charitable than negative *moral* attributions. From the attributor's perspective, calling someone stupid typically seems less harsh than calling them morally bad. Insofar as the Razor can shift people's construal of conflict in that way, it can reduce the conflict's severity and let them treat their opponents better. Or so goes one line of support for premise (2).

An important question is whether merely thinking about the Razor is sufficient to reframe conflict charitably. Why couldn't someone reflect on the Razor's advice but subsequently see their conflict no differently, or even see it more fractiously and tendentiously? That may happen occasionally, though we expect otherwise in some important cases.

What follows is a speculative idea in favor of that expectation. The Razor is intuitively resonant—nobody needs to argue for its *prima facie* plausibility. Why is that? Just consider the experience of reflecting on the principle in the heat of conflict. The Razor invites you to swap out malice attributions for stupidity attributions. What we called the "Gestalt shift" allows you to see others' mistakes as dumb, not evil. Switching one type of attribution for another may in part underwrite the Razor's intuitive appeal, we speculate. When people exchange malice for stupidity attributions, they may *feel good* about themselves—literally. Imagine you are in a dispute and you think to yourself, "My opponents are evil!" Then you reflect on the Razor and come to think, "No, my opponents are just dumb!" That shift casts you in a positive light, not your opponent.

That difference is crucial. One way that your stupidity attributions could improve a conflict is by making you feel more positively about your opponents. Alternatively, your stupidity attributions could improve a conflict by making you view *yourself* more positively. We suspect the latter is what might happen. When using the Razor, you may feel flattered—seeing yourself as smarter than your opponents, instead of judgmentally scolding their inferior moral character. Pat yourself on the back for being so charitable; climb atop on your high horse for exhibiting civility and maturity. And insofar as the Razor can convey to users the idea that they are in some way *better*—more intelligent, reasonable, charitable, or virtuous—for seeing their opponents' errors flowing from stupidity than from malice, there is some reason to expect the Razor can downgrade the severity of conflict. Potentially, the principle helps people to act nicer toward others by making them feel nicer about themselves.[8]

But even if this represents one viable path for the Razor to improve relationships between opponents, we think premise (2) requires considerably more support. For a wide range of situations, we don't know whether the Razor works as suggested. Premise (2) implies empirical predictions that are far from certain. Although none of this can be settled from the

[8] Alternatively, the Razor could sometimes make people puffed up about themselves and thus less likely to act charitably toward their perceived-to-be-stupid opponents.

armchair, we offer five observations that reveal why it's hard to predict how the Razor influences thinking.

First, when intelligence or epistemic competence is valued by attributors more highly than good motives or moral character, attributions of stupidity could amplify conflict more than attributions of malice. Malice is often a greater insult than stupidity, but there are exceptions. Imagine a successful scientist who suffers from imposter syndrome (Langford and Clance 1993). Conceivably, this scientist could assume it's worse to be called stupid than morally bad, given she prizes intelligence above character. How the Razor works in such cases is uncertain.

Second, even though the Razor can in principle moderate overly harsh attributions, people could find it difficult to follow the Razor because malice attributions are so compelling phenomenologically. Withholding our negative moral attributions may feel wrong—*Those evil bastards deserve harsh judgment!* Indeed, one key insight from recent moral and political psychology is that partisans do not play nice (Skitka et al. 2021). The dynamics of negative moral attribution can be illustrated using metaphors of war and violence. Partisans brandish a Karmic Bazooka and try to blast their enemies into submission—*Take this, you deplorable monsters!* They viscerally feel that their judgments are righteous, just, and true. Prompting people to reflect seriously on their opponents' epistemic limitations may be tricky when they feel their opponents are immoral. And when they feel personal grievance, they may be unable to overlook perceived injustice (Ditto and Rodriguez 2021). In general, morality can bias and override the kind of epistemic evaluation the Razor encourages. So, we think there are important questions whether, and to what extent, the principle can intervene in conflicts fueled by moral convictions and visceral emotions. Potentially, it won't help precisely where it is needed most.[9]

Third, when people posit an epistemic defect to explain an opponent's error, they may be unable to make sense of the existence of that defect in non-moral terms. They will think someone's stupidity is morally culpable or otherwise a sign of a moral flaw. In other words, people can believe their opponent errs because of stupidity and yet come to believe he's stupid because he's bad. The Razor requires people to presuppose some epistemic defect is not a moral defect, but we're suggesting there can be inadvertent "spillover" from epistemic defect to moral defect. For example, during the Covid-19 pandemic many so-called anti-maskers claimed that medical-grade face masks do not protect against transmission of the coronavirus. Explaining an anti-masker's error using epistemic defects is easy, but an observer might feel the error is so grievous and so easily fixable that the anti-masker must be morally culpable. In such cases, the Razor could reinforce malice attributions by making them seem to follow from facts about stupidity.

Fourth, the Razor may make people unsure how to interpret their opponents. Their framing of a conflict may fall somewhere in between charitable and harsh. What happens then? One

[9] A related point concerns the personality traits needed to deploy the Razor. If following the principle calls for some degree of intellectual humility (Ballantyne forthcoming-a), many people who would benefit from it won't deploy it. (Thanks to Daniel Relihan for discussion.)

possibility is suggested in studies by Steven Fein and James Hilton, who examined subjects' suspicion about an actor's motivations (Fein and Hilton 1994; Fein, Hilton, and Miller 1990). For Fein and Hilton, *suspicion* is "a state in which perceivers actively entertain multiple, plausibly rival, hypotheses about the motives or genuineness of a person's behavior" (1994, 168). They found that suspicion makes people less inclined to accept others' word at face value and to see them as less likeable (1994). Insofar as the Razor induces suspicion, it could harden or even intensify conflict.[10] That sort of possibility challenges the idea that the Razor typically makes people get along better with opponents.[11]

A final point will be familiar by now. The Razor can have a "dark side," because being too charitable is dangerous. Failing to recognize that other people have bad motives can turn someone into an easy target for abuse. To take an example from politics, bipartisanship helps a group reach important compromises, but these efforts often require people to downplay the possibility that their opponents have bad motives or even autocratic ambitions. Misjudging malicious opponents is perilous—the word "appeasement" is one troubling reminder. But the challenges here aren't only because of our overly charitable judgments concerning opponents. Our opponents, seeing we are open to extra-generous interpretations of them, may try to manipulate us. They may judge that *we* are stupid. Recall that the Razor's relevance in any situation hinges on the subjective plausibility of stupidity and malice attributions. If stupidity is a non-starter as an explanation for our opponents' errors, they could self-present in ways that raise the plausibility of their stupidity, thus trying to lower the plausibility of their malice. Our opponents could hoodwink us by managing our impressions. The deceptive self-presentation points toward a cynical maxim: *Never admit to malice when stupidity will suffice.* People often want others to give them the benefit of the doubt about malice and so they may sacrifice their perceived epistemic competence to save their moral reputation—*"I just didn't know..."* In short, if we use the Razor, bad actors may take advantage of our good will and exploit us.

In our discussion of the possibility that the Razor makes people more charitable, we assumed it can be good advice even if it doesn't tend to improve the accuracy of moral attributions. But notice how it could secure different epistemic benefits when it moderates people's harsh judgments about their opponents. Suppose you initially frame a dispute as caused by your opponent's malice. Then the Razor prompts you to see your opponent's mistake as due to stupidity. That new interpretation makes you more open to productive dialogue and

[10] Fein and Hilton note: "Suspicious perceivers may not be able or care to hide their suspicions from the target of their suspicions. They may act in a more cold and distant way, perhaps guarding themselves from any self-disclosures that the target potentially could use to his or her advantage. This set of behaviors is likely to create a vicious circle in which the perceivers' suspicions lead the target to behave in strange ways, thus reinforcing the perceivers' suspicions" (1994, 193).

[11] Our worry is that premises (1) and (2) could easily fail to count as good evidence for (3). That is, even if seeing conflict in terms of divergent motives makes people get along less well [i.e., (1)] and following the Razor tends to prevent someone from seeing conflict in such terms [i.e., (2)], the Razor could still increase someone's harshness toward her opponents because it makes her suspicious.

cooperation. Rather than dismissing your opponent as a moral miscreant, you try to share persuasive evidence.[12] You show greater willingness to listen to your opponent's reasoning, at least while adjusting your angle of dialectical attack. Your greater openness to engagement and listening could bring epistemic benefits—not just for your opponent but for you as well.[13]

So far, our exploration of the Razor has focused on matters of interpersonal judgment. But we think the Razor obscures an important matter: the person using the Razor.

5. Look in the mirror

The Razor is supposed to aid our judgment by prompting us to think about other people. It fixes our attention on our judgment of them, not ourselves or our judgment of ourselves. Even if the Razor sometimes turns out to be good advice in the sense that it yields benefits, it can easily fail to deliver when we lack knowledge to deploy it properly. We will note two limitations of self-insight—specifically, insights into one's own knowledge—that impede the principle's operation. The first is a type of "blind spot" that prevents us from recognizing the ways in which others' belief-forming processes lead to mistakes; the second is a failure to scrutinize weaknesses in our own thinking and then using the Razor to bolster our erroneous confidence.

As we begin to set out the "blind spot" idea, consider a basic observation: people sometimes project their knowledge onto others. They assume others know what they themselves know and feel what they feel (Ditto and Koleva 2011; Camerer, Loewenstein, and Weber 1989). For instance, suppose you believe some claim is obviously true and compelling to everyone in your community. When your neighbor rejects the claim, you may reactively explain his deviation by attributing malice. He should know better—he denies something you believe is obvious to him. Of course, if your projection is mistaken, your attribution of malice may be too hasty.

Move from the example to some more general ideas. People in conflicts make assumptions concerning their opponents' belief-forming methods or processes. Opponents' methods are not directly observable but must be inferred from behavior or based on pieces of background knowledge. As we already noted, people may project their own situation onto others by assuming self-similarity: their opponents form beliefs by using the same kind of processes they themselves use. But assumptions of similarity can fail because others' thinking may go in directions people do not and perhaps cannot anticipate. Take some examples. You assume someone's political opinions are based on reports from one pundit; in fact, his opinions are based on reports from another pundit. You take for granted that a person formed a belief

[12] We speculate that people tend to think moral change is often harder, and less likely achieved, than "epistemic change." That is, becoming a more well-informed or accurate thinker tends to be seen as easier than becoming a morally better person.

[13] Of course, someone's being overly charitable could allow her opponents to lead her away from the truth when she is correct. Supposing she is more likely to engage with opponents she regards as stupid than evil, her opponents get opportunities to mislead her. (Thanks to Peter Seipel for discussion.)

through testimony; instead, her belief is based on perceptual experience. Or perhaps you presume a claim is intuitively obvious to someone; but what's obvious to him is the negation of that claim. In such situations, you find yourself in a *method blind spot*: you neither recognize how others formed their beliefs nor the particular epistemic defects that give rise to their mistakes. You are oblivious to why they think as they do. A method blind spot could be sizable when you disagree with people whose ideological commitments or cultural backgrounds are quite different than yours. In general, biases on perspective-taking—such as the "curse of knowledge" (Camerer, Loewenstein, and Weber 1989), the inability to think about a topic from a less informed perspective—suggest that someone's high confidence in her own views may lead to confusions about what and how their opponents know.

What exactly are the implications of the method blind spot for using the Razor? The issues here are complex, but we'll note one important type of situation. Prompted by the Razor, you may wonder *ad nauseam* whether your opponent might have erred due to stupidity; but if his epistemic defects are hidden inside your blind spot, you won't see them. You may too readily conclude he's morally bad, not stupid, even in cases where his error is due to stupidity. Deploying the Razor inside a blind spot makes matters worse.

To be sure, all of us frequently find ourselves in method blind spots. The best advice to break out is straightforward. Talk to your opponents. Try to find out what they know and don't know, and how they arrived at their opinions. Seek to better understand their motives. Trying to suss out someone's belief-forming methods is not always easy or possible, but engagement is often worth the effort.[14] In support of this advice, some research suggests that sharing facts doesn't bridge political divides as effectively as sharing personal experience does. In one study, subjects had greater respect for their opponents' moral beliefs when their opponents appealed to personal experiences, not facts (Kubin et al. 2021). What may be going on is this: when someone shares her experiences, that helps to illuminate the method she uses to form opinions, allowing her ideological opponents to interpret her opinions more charitably. Crucially, efforts to eliminate or reduce a method blind spot require our awareness that there is, or could be, one. We can't use the Razor well if we are unknowingly trapped in the blind spot and can't distinguish someone's stupidity from malice.

Turn to a second failure of self-insight: the Razor can fuel confidence in mistaken views. As we noted, the principle takes your own correctness as a presumed "fixed point" for reasoning and tries to pivot your thinking about your opponent. But what happens if your opponent is right and you are wrong? Or if both of you are wrong? And what if you are the one who needs to recognize your own stupidity or malice? Even granting you have reliable access to facts concerning the goodness or badness of your intentions and moral character, some epistemic defects can be tricky to identify: you may lack evidence or conceptual grasp to recognize your ignorance and spot your flaws (Dunning 2011; Ballantyne forthcoming-c). But if you are oblivious to facts indicating you are wrong, the Razor could become a source of misguided confidence.

[14] Ballantyne (forthcoming-b) discusses challenges of engaging with opponents when our knowledge about them is limited and conflict is highly polarized.

To see why, notice that appealing to an intuitively resonant principle may encourage you to believe you are being reasonable while working through a conflict. The act of consulting the principle seems like what a good thinker would do. But aided by the Razor your reflection could easily focus too much on your opponent and too little on yourself. That would not be surprising given the ubiquitous confirmation bias (Nickerson 1998). People are more adept at finding flaws in arguments that threaten their views than flaws in supporting arguments, all things being equal (Ditto and Lopez 1992; Ditto 2009). A similar effect produces asymmetrical levels of scrutiny toward your opponents' intellectual and moral character compared to your own character. You pick apart your opponents to uncover their stupidity or malice while giving yourself a free pass. Sometimes, when you are mistaken, the Razor may pull your focus away from where it should be while simultaneously making you feel you've done your level best.

To dodge that problem, we could supplement the Razor with another rule: *Never attribute someone's error to either a moral or epistemic defect before you have seriously investigated your own moral and epistemic competence*. The idea is reminiscent of age-old wisdom: "First take the log out of your own eye, and then you will see clearly to take the speck out of your brother's eye." Even if the supplementary rule can prevent the Razor from inflating confidence in wrong views, it restricts the Razor's scope. Moreover, for many controversial issues, knowing you are intellectually competent turns out to be challenging (Ballantyne 2019).

All of this raises questions concerning the Razor's proper use. No principle can resolve every problem, and so wielding the Razor without coordination with other principles can lead to trouble. People have a toolbox of principles and rules for judgment and, ideally, they use the right ones at the right times. Our point is that anyone trying to deploy the Razor should be wary of two mistakes: using the principle inside a method blind spot or having it boost confidence in erroneous views. What is needed is greater self-awareness. When using any kind of razor, it's wise to use a mirror.

6. Conclusion

Is Hanlon's Razor good or bad advice? In this essay, we criticized two proposals in favor of the Razor. One sees the benefits of the principle in terms of making us more accurate. The other sees benefits in terms of making us more charitable. Our discussion has been preliminary, but we hope careful empirical investigation can illuminate when and why the Razor is beneficial, if it is. For the time being, what else can we say about this principle?

The Razor attempts to address the problem of detecting facts that explain opponents' mistakes. Why do our opponents screw up? For hypermoralists, detecting stupidity in the noise of malice can be difficult: we are too eager to attribute bad motives and unsavory character to people who disagree with us. When we try to explain their mistakes, we are subject to two distinct errors:

> *Misidentifying-stupidity error*: attributing an error to malice that is due to stupidity
>
> *Misidentifying-malice error*: attributing an error to stupidity that is due to malice

The idea driving the Razor is simple enough. People make misidentifying-stupidity errors too frequently and they should minimize those errors by risking misidentifying-malice errors. The Razor attempts to adjust our criterion for detecting the source of opponents' mistakes. People should see stupidity more often in their opponents, even if that means they sometimes see stupidity where there is in fact malice.

Regrettably, the Razor is not sensitive to what matters most. As we pointed out earlier, one danger of misidentifying-malice errors is exploitation by bad actors. Ruminating on the possibility of abuse, or getting burned by a bad actor, may lead people to adjust their criterion in the opposite direction of the Razor's advice, seeing others as being bad more than stupid. Some people may feel that being highly sensitive to malice is a safe policy. They will not be exploited by opponents if they assume each and every opponent is bad. Of course, the costs of such a strategy are enormous—entrenched conflict, negativity, and polarization. Even if people who see malice everywhere become "safe" in some sense, they will find themselves trapped in a world of profound negativity. And given the costs of such an outlook, people may feel the Razor was a good idea after all: viewing our opponents as stupid more than bad makes our interactions more positive and allows compromise and cooperation.

We are not required to choose, once and for all, between a "safe" but negative criterion and a more open, cooperative one. People adjust their criterion for detecting stupidity or malice on the fly. Worried about the costs of exploitation, we crank up our sensitivity to moral defects; and worried about the costs of failing to cooperate and compromise, we boost our sensitivity to epistemic defects. In fine-tuning our criterion, circumstances matter. Let us illustrate the point using a pair of examples.

In 2015, U.S. Vice President Joseph Biden delivered an address to the graduating class at Yale University. Biden recounted a story from his time as a Senator back in the 1970s. One Republican Senator, Jesse Helms, had argued vigorously against legislation that was the precursor of the Americans with Disabilities Act. Biden was disgusted. He recalled meeting with a Democratic colleague, Majority Leader Mike Mansfield, and complaining that Helms had "no social redeeming value" and didn't care about disabled people. In response, Mansfield shared a story, which Biden recounted to his audience at Yale:

> [T]hree years earlier, Jesse and Dot Helms, sitting in their living room in early December before Christmas, reading an ad in the Raleigh Observer, the picture of a young man, 14-years-old with braces on his legs up to both hips, saying, all I want is someone to love me and adopt me. [Mansfield] looked at me and he said: and they adopted him, Joe.
>
> I felt like a fool. [Mansfield] then went on to say: Joe, it's always appropriate to question another man's judgment, but never appropriate to question his motives because you simply don't know his motives. (Biden 2015)

As Biden tells the story, Mansfield's advice became an insight for Biden's work in government over the decades. Biden adjusted his criterion for viewing his colleagues' errors: "[W]hen you question a man's motive… it's awful [sic] hard to reach consensus. It's awful

[sic] hard having to reach across the table and shake hands. No matter how bitterly you disagree, though, it is always possible if you question judgment and not motive… Resist the temptation to ascribe motive" (Biden 2015).

Some may suspect Biden's good will toward others makes him an easy target for abuse by malicious actors.[15] With that in mind, consider a contrasting example from a classic essay, "How to Swim with Sharks: A Primer" (Cousteau 1973). The essay offers darkly humorous advice for dealing with dangerous people ("sharks"):

> Assume all unidentified fish are sharks. Not all sharks look like sharks, and some fish that are not sharks sometimes act like sharks. Unless you have witnessed docile behavior in the presence of shed blood on more than one occasion, it is best to assume an unknown species is a shark. Inexperienced swimmers have been badly mangled by assuming that docile behavior in the absence of blood indicates that the fish is not a shark. (Cousteau 1973, 525–26)

Let's suppose each of us must lock in our criteria for detecting malice and the options are to be like Biden or the swimmer. Which option would you choose? Of course, these are extreme positions—never assuming malice or always assuming it—and both have genuine downsides. We should instead hope to find criteria that hit a "sweet spot" between the extremes.

What this little thought experiment exposes is a fundamental limitation for any automatic adjustment of our criteria, whether the adjustment tips the scales toward stupidity or malice. An automatic adjustment won't give people a chance to reap the benefits of cooperation and avoid the costs of abuse. The goal should be to become more sensitive to the difference between malice and stupidity. We need to discriminate between errors caused by moral defects and epistemic defects.

Does that mean the Razor is bad advice? Not necessarily. We think the Razor encourages users to find the "sweet spot." When an error is *adequately explained* by stupidity, says the Razor explicitly, attribute stupidity. The Razor adds in an implicit exhortation, too: whether or not stupidity provides an adequate explanation in any situation deserves careful scrutiny and then and only then should explanations based on malice come into play. Wise judgment is vital. But we are on our own: the Razor express a lofty aspiration without explaining how to attain it.

In a world of conflict and partisanship, it's tempting to believe that stepping back from sectarian battles between "good and evil" and considering more mundane explanations for

[15] Importantly, some people won't accept at face value Biden's proposal to leave motives unquestioned. Think about how his proposal would be viewed inside the U.S. Senate. Imagine Biden's fellow senators wondering privately: "Is Joe hiding relevant motives? What doesn't he want us to know when we hear his arguments for legislation? He insists he won't question our motives and wants us to do likewise, but could that be a ploy to soften us up?" (No critic will be moved by a Bidenesque rejoinder: "C'mon, man!")

our opponents' behavior could infuse greater humility and civility into discourse and debate. There may be some truth to that. But we have tried to show how judgment in the real world gets complicated: our knowledge of others' thinking and motives is limited, the dynamics of social judgment are complex, and people sometimes act in bad faith. The challenge of becoming more discriminating about other people requires more than a clever, resonant principle like the Razor. And it doesn't help to wield that principle incautiously, without subtle judgment or sensitivity to context. Hanlon's Razor is not a machete for judging others. It's more like a scalpel. We should handle it with care.[16]

Bibliography

Alicke, Mark D. 2000. "Culpable Control and the Psychology of Blame." *Psychological Bulletin* 126 (4): 556–74.

Ballantyne, Nathan. 2016. "Verbal Disagreements and Philosophical Scepticism." *Australasian Journal of Philosophy* 94 (4): 752–765.

Ballantyne, Nathan. 2019. *Knowing Our Limits.* New York: Oxford University Press.

Ballantyne, Nathan. Forthcoming-a. "Recent Work on Intellectual Humility: A Philosopher's Perspective." *Journal of Positive Psychology*. doi: 10.1080/17439760.2021.1940252.

Ballantyne, Nathan. Forthcoming-b. "The Fog of Debate." *Social Philosophy and Policy*.

Ballantyne, Nathan. Forthcoming-c. "Tragic Flaws." *Journal of the American Philosophical Association*. doi: 10.1017/apa.2020.39.

Biden, Joseph. 2015. "Remarks by the Vice President at Yale University Class Day." https://obamawhitehouse.archives.gov/the-press-office/2015/05/17/remarks-vice-president-yale-university-class-day

Böhm, Robert, Isabel Thielmann, and Benjamin E. Hilbig. 2018. "The Brighter the Light, the Deeper the Shadow: Morality Also Fuels Aggression, Conflict, and Violence." *Behavioral and Brain Sciences* 41: e98. doi:10.1017/S0140525X18000031.

Boyd, Robert, and Peter J. Richerson. 1992. "Punishment allows the evolution of cooperation (or anything else) in sizable groups." *Ethology and Sociobiology* 13 (3): 171–195.

[16] For helpful comments and conversations, both authors would like to acknowledge Matthew Ballantyne, Andrew Bailey, Jared Celniker, Pasha Dashtgard, Noah Hahn, Victor Kumar, Mertcan Gungor, Cathal O'Madagain, Daniel Relihan, Peter Seipel, Shiri Spitz, Joel Walmsley, and Benjamin Wilson. Thanks to members of the Hot Cognition Lab at UC Irvine for discussion in August 2021. NB would like to express his gratitude to the John Templeton Foundation (grant 61014) for generous support.

Camerer, Colin, George Loewenstein, and Martin Weber. 1989. "The Curse of Knowledge in Economic Settings: An Experimental Analysis." *Journal of Political Economy* 97 (5): 1232–1254.

Chalmers, David J. 2011. "Verbal Disputes." *The Philosophical Review* 120 (4): 515–66.

Cheek, Nathan N., and Emily Pronin. Forthcoming. "I'm Right, You're Biased: How We Understand Ourselves and Others." In *Reason, Bias, and Inquiry: New Perspectives from the Crossroads of Epistemology and Psychology*, edited by Nathan Ballantyne and David Dunning. New York: Oxford University Press.

Clark, Cory J., Jamie B. Luguri, Peter H. Ditto, Joshua Knobe, Azim F. Shariff, and Roy F. Baumeister. 2014. "Free to Punish: A Motivated Account of Free Will Belief." *Journal of Personality and Social Psychology* 106 (4): 501–13.

Cousteau, Voltaire. 1973. "How to Swim with Sharks: A Primer." *Perspectives in Biology and Medicine* 16 (4): 525–528.

Ditto, Peter H., and David F. Lopez. 1992. "Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions." *Journal of Personality and Social Psychology* 63 (4): 568–584.

Ditto, Peter H. 2009. "Passion, Reason, and Necessity: A Quantity-of-Processing View of Motivated Reasoning." In *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation*, edited by Tim Bayne and Jordi Fernández, 23–53. New York: Psychology Press.

Ditto, Peter H., and Spassena P. Koleva. 2011. "Moral Empathy Gaps and the American Culture War." *Emotion Review* 3 (3): 331–332.

Ditto, Peter H., and Cristian G. Rodriguez. 2021. "Populism and the Social Psychology of Grievance." In *The Psychology of Populism: The Tribal Challenge to Liberal Democracy*, edited by Joseph P. Forgas, William D. Crano, and Klaus Fiedler, 23–41. New York: Routledge.

Dunning, David. 2011. "The Dunning-Kruger effect: On being ignorant of one's own ignorance." In *Advances in Experimental Social Psychology*, Volume 44, edited by James Olson and Mark P. Zanna, 247–296. New York: Elsevier.

Fehr, Ernst, and Simon Gächter. 2002. "Altruistic Punishment in Humans." *Nature* 415: 137–140.

Fein, Steven, and Hilton, James L. 1994. "Judging Others in the Shadow of Suspicion." *Motivation and Emotion* 18 (2): 167–98.

Fein, Steven, James L. Hilton, and Dale T. Miller. 1990. "Suspicion of Ulterior Motivation and the Correspondence Bias." *Journal of Personality and Social Psychology* 58 (5): 753–64.

Feltz, Adam. 2007. "The Knobe Effect: A Brief Overview." *The Journal of Mind and Behavior* 28 (3/4): 265-77.

Friedman, Jane. 2013. "Suspended Judgment." *Philosophical Studies* 162 (2): 165–81.

Henrich, Joseph, Richard Mcelreath, Abigail Barr, Jean Ensminger, Clark Barrett, Er Bolyanatz, Juan Camilo Cardenas, Michael Gurven, and Edwins Gwako. 2006. "Costly Punishment Across Human Societies." *Science* 312: 1767–1770.

Kahneman, Daniel. 2003. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist* 58 (9): 697–720.

Knobe, Joshua. 2010. "Person as Scientist, Person as Moralist." *Behavioral and Brain Sciences* 33 (4): 315–29.

Kubin, Emily, Curtis Puryear, Chelsea Schein, and Kurt Gray. 2021. "Personal experiences bridge moral and political divides better than facts." *Proceedings of the National Academy of Sciences* 118 (6): e2008389118. doi: 10.1073/pnas.2008389118

Langford, Joe, and Pauline Rose Clance. 1993. "The Imposter Phenomenon: Recent Research Findings Regarding Dynamics, Personality and Family Patterns, and Their Implications for Treatment." *Psychotherapy* 30 (3): 495–501.

La Rochefoucauld. 1678/2007. *Collected Maxims and Other Reflections*, translated with an introduction and notes by E. H. and A.M. Blackmore and Francine Giguère. Oxford: Oxford University Press.

Leslie, Alan M., Joshua Knobe, and Adam Cohen. 2006. "Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment." *Psychological Science* 17 (5): 421–27.

Nickerson, Raymond S. 1998. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises." *Review of General Psychology* 2 (2): 175–220.

Quote Investigator. 2016. "Never Attribute to Malice That Which Is Adequately Explained by Stupidity." Last modified 30 December 2016. https://quoteinvestigator.com/2016/12/30/not-malice/.

Rozin, Paul. 1999. "The Process of Moralization." *Psychological Science* 10 (3): 218–21.

Schwarz, Norbert, Lawrence J. Sanna, Ian Skurnik, and Carolyn Yoon. 2007. "Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns." *Advances in Experimental Social Psychology* 39: 127–61.

Skitka, Linda J., James Hou-fu Liu, Yiyin Yang, Hui Chen, Li Liu, and Lun Xu. 2013. "Exploring the Cross-Cultural Generalizability and Scope of Morally Motivated Intolerance." *Social Psychological and Personality Science* 4 (3): 324–31.

Skitka, Linda J., Christopher W. Bauman, and Edward G. Sargis. "Moral Conviction: Another Contributor to Attitude Strength or Something More?" *Journal of Personality and Social Psychology* 88 (6): 895–917.

Tetlock, Philip E. 2003. "Thinking the unthinkable: sacred values and taboo cognitions." *Trends in Cognitive Sciences* 7 (7): 320–24.

Tetlock, Philip E., Orie V. Kristel, S. Beth Elson, Melanie C. Green, and Jennifer S. Lerner. 2000. "The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals." *Journal of Personality and Social Psychology* 78 (5): 853–70.

Walmsley, Joel, and Cathal O'Madagain. 2020. "The Worst-Motive Fallacy: A Negativity Bias in Motive Attribution." *Psychological Science* 31 (11): 1430–38.

Waytz, Adam, Liane L. Young, and Jeremy Ginges. 2014. "Motive attribution asymmetry for love vs. hate drives intractable conflict." *Proceedings of the National Academy of Sciences* 111 (44): 15687–92.

Wilson, Timothy D., and Nancy Brekke. 1994. "Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations." *Psychological Bulletin* 116 (1): 117–142.

Wilson, Timothy D., David B. Centerbar, and Nancy Brekke. 2002. "Mental Contamination and the Debiasing Problem." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by Thomas Gilovich, Dale Griffin, and Daniel Kahneman, 185–200. Cambridge: Cambridge University Press.

Zaal, Maarten P., Rim Saab, Kerry O'Brien, Carla Jeffries, Manuela Barreto, and Colette van Laar. 2017. "You're either with us or against us! Moral conviction determines how the politicized distinguish friend from foe." *Group Processes and Intergroup Relations* 20 (4): 519–39.