C15

CHAPTER 15

REDIRECTING RAWLSIAN REASONING TOWARD THE GREATER GOOD

JOSHUA D. GREENE, KAREN HUANG, AND MAX BAZERMAN

C15.S1

15.1 Introduction

C15.P1

At the heart of John Rawls's masterwork, *A Theory of Justice*, is a thought experiment. Rawls asks: What kind of a society would we choose if we didn't know who in that society we would be? The question is hypothetical, but the aim is to inform our thinking about the real world. A just society, Rawls argues, is one that we would choose if we were unbiased. It is, more specifically, one we'd choose if we lacked the information necessary to tilt the scales of justice toward our individual interests. The decision-makers in this thought experiment are said to be in the 'Original Position', and their choice is made from behind a 'Veil of Ignorance' (VOI). As Rawls (1971: 12) explains:

C15.P2

> Among the essential features of this situation is that no one knows his place in society, his class position or social status, nor does anyone know his fortune in the distribution of natural assets and abilities, his intelligence, strength and the like. I shall even assume that the parties do not know their conceptions of the good or their special psychological propensities.

C15.P3

One could add that the decision-makers are likewise ignorant of their races, cultural backgrounds, gender identities, sexual orientations, and so on. The key idea, once again, is that the decision-makers lack the knowledge needed to bias their decisions, for example, by choosing principles that favour men over women, one race over another, etc. Although the ultimate goal is to illuminate the principles of justice, the decision-makers are assumed to be purely self-interested, as well as rational. In the Original Position, the absence of bias comes not from the virtue of the decision-makers, but from the structure of the decision. It resembles the 'I cut, you choose' method for cutting a cake,[1] a procedure that turns

[1] The cases we'll consider, unlike the cake-cutting case, are ones of 'pure procedural justice' (Rawls 1971: 85–6), with no independent criterion for fairness.

selfish choices into fair outcomes. Likewise, says Rawls, selfish individuals who choose their governing principles from behind a VOI, with all biasing information withheld, will choose a just set of principles.

C15.P4 Rawls applied this thought experiment to the most fundamental question of political philosophy: According to what principles should a society be organized? The same logic, however, can be applied to more specific moral dilemmas, and by ordinary people. In two recent sets of experiments, we have done just this (Huang, Greene, and Bazerman 2019; Huang, Bernhard, Barak-Corren, Bazerman, and Greene 2021).[2] Here we summarize our main experimental results and consider their implications. First, we argue that our findings provide further support for consequentialist approaches to ethics. Second, and more importantly, we argue that veil-of-ignorance reasoning may be a useful and underappreciated tool for thinking about real-world moral problems. We highlight the ability of VOI reasoning to foster more impartial decision-making and promote the greater good across a variety of domains, from the ethics of self-driving cars to healthcare to charitable giving.

C15.S2

15.2 The veil of ignorance and moral dilemmas

C15.P5 We'll begin, as we must, with the footbridge dilemma (Thomson 1985). For the uninitiated: a runaway trolley is headed toward five people. You and the large man (or man with a large backpack) are on a footbridge spanning the tracks. If you do nothing, the five will die. But you can save the five by pushing the man off the footbridge and onto the tracks, killing the man but blocking the trolley and thus saving the five. (Yes, this will work, and no, you cannot sacrifice yourself because you are too light to stop the trolley.) Is it morally acceptable to push?

C15.P6 The argument in favour of pushing is a straightforward consequentialist/utilitarian[3] one: Pushing will save more lives. Nevertheless, most people say that it's wrong to push, even under the assumption that it will save more lives. The argument against pushing is typically framed in deontological terms: 'The ends don't justify the means' or 'Pushing the man to his death would violate his rights' (Thomson 1990).

C15.P7 What happens, though, if we engage in a bit of veil-of-ignorance reasoning about this case? (See also Hare 2016 for an earlier use of this approach.) There are six people who are unambiguously affected by the decision of whether or not to push: the pushable man on the footbridge and the five people on the tracks who could be saved by pushing. Suppose that you have an equal probability[4] of being each of these six people. From a purely self-interested

[2] For earlier empirical uses of VOI reasoning, applied to Rawls's original questions concerning society's organizing principles, see Frohlich, Oppenheimer, and Eavey (1987); Frohlich and Oppenheimer (1993).

[3] Although utilitarianism is a special case of consequentialism, these philosophical terms are, for present purposes, interchangeable, given certain reasonable assumptions about the hedonic consequences of, and moral significance of, more vs fewer deaths.

[4] In all of our experiments we assume that one has an equal probability of being each of the people affected by the decision. Rawls, in his original VOI thought experiment, assumes that the probabilities are unknown. Following Harsanyi (1955; 1975), we use an equiprobability assumption because (in

perspective, what would you want the decision-maker to do? The answer seems clear. You would want the decision-maker to push, as you would rather have a 5 in 6 chance of living than a 1 in 6 chance of living.

C15.P8 But what moral implications, if any, does this have? After all, the original footbridge dilemma poses a moral question, while the VOI version asks for a self-interested preference about a situation even more bizarre than the original. The same Rawlsian logic applies. According to Rawls, a just social order is one that you would choose (selfishly) if you didn't know which position in that social order you would occupy. So, why not make an analogous argument here? If you didn't know who in this situation you were going to be, you would want the decision-maker to push. So, why not say that pushing to save five lives is the more just thing to do?

C15.P9 At this point, some readers may be relieved to hear that we are not going to apply this analogy in the reverse direction, arguing that VOI reasoning underwrites a general utilitarian social philosophy, as claimed by Harsanyi (1955; 1975). Some of us happen to think that Harsnayi was right about this, but we will not press that case here. Instead, we are only claiming that veil-of-ignorance reasoning favours the greater good across a range of specific dilemmas, including some with real-world significance.

C15.P10 We have observed such effects across two sets of experiments (Huang, Greene, and Bazerman 2019; Huang, Bernhard, Barak-Corren, Bazerman, and Greene 2021). This holds not only for the classic footbridge dilemma, but also for a range of more realistic cases, as explained below. To be clear, our finding is not simply that people give more utilitarian answers to the VOI versions of these dilemmas. Rather, it's that thinking through the VOI version of a dilemma changes the way people respond to the standard version of that dilemma. For example, in response to the VOI footbridge case, a typical participant will conclude that she would want the decision-maker to push if she had an equal chance of being each of the six people affected by the decision. But then, when she subsequently considers the standard footbridge case, she's more likely to say it's morally acceptable to push. This two-step process mirrors Rawls's use of VOI reasoning in *A Theory of Justice*, whereby the purpose of the VOI thought experiment is to inform our subsequent thinking about the original moral question.

C15.P11 Generating approval for pushing people off footbridges may not seem like a worthy accomplishment, but that was not our goal. Most of our experiments addressed more realistic decisions, and ones for which the utilitarian option, while controversial, is more morally palatable and easier to take seriously than the proverbial footbridge push. In one of the experiments reported in our first paper (Huang, Greene, and Bazerman 2019), participants considered a bioethical dilemma adapted from Robichaud (2015) involving the provision of oxygen during the aftermath of an earthquake (Robichaud's original dilemma focused on a

addition to its being simpler) we believe that it more faithfully adheres to the purpose of the VOI thought experiment as a device for encouraging more impartial thinking. If the idea is to give an unbiased answer, one that gives equal weight to each person, then why not give oneself an equal probability of being each person? As one of us has argued elsewhere (Greene 2013: 383–), we suspect that Rawls's decision to make the odds unknown rather than equal is actually a fudge factor. His use of unknown odds makes extreme risk aversion in the Original Position seem more plausible, and this in turn helps make Rawls's favoured 'maximin' rule seem more plausible.

terrorist attack.) Engaging in VOI reasoning about this dilemma led people to favour using the oxygen in a way that would save more lives. In other experiments we used a dilemma concerning the ethics of autonomous vehicles (AVs), adapted from Bonnefon et al. (2016). Here, an AV is headed toward several pedestrians who will be killed if it stays on course. The AV can avoid killing these people by swerving, but this will send it into a concrete wall and kill the AV's single passenger. In the VOI version of this dilemma, people typically say that they would want the car to swerve if they knew they would have an equal chance of being each of the people affected by the decision. And then, after considering the VOI version of the AV dilemma, people were more likely to endorse a policy that would require AVs to minimize the total loss of life, even at the expense of AV passengers. In one of these experiments, VOI reasoning resulted in 83 per cent of participants' approving of the utilitarian AV policy, as compared to 58 per cent in the control condition, turning a highly controversial proposal into one with fairly strong consensus.

C15.P12 In another experiment in this series, we examined the effect of VOI reasoning on real donation decisions. Participants in the US were presented with descriptions of two real charities, one in the US and one in India, both of which fund procedures that restore people's vision. The Indian charity, however, is more effective because the same amount of money can help twice as many people. Running this dilemma through the VOI, one can imagine having a 1 in 3 chance of being helped by a donation to the US charity and a 2 in 3 chance of being helped if the money, instead, goes to the Indian charity. As predicted, thinking through the VOI version of this dilemma made people more likely to direct a real donation to the more effective charity.

C15.P13 In a second set of experiments (Huang, Bernhard, Barak-Corren, Bazerman, and Greene 2021), we applied VOI reasoning to the ventilator dilemma faced by doctors in Italy (and elsewhere) during the early phases of the COVID-19 crisis (Mounk 2020). We focused, more specifically, on the question of whether age should be a factor in the allocation of life-saving resources. Under ordinary circumstances, medical resources are allocated under a 'first come, first served' rule. This is generally considered a fair principle, as it does not discriminate on the basis of patients' personal characteristics such as wealth, race, gender, religion, or age. However, under conditions of scarcity, there is a utilitarian argument favouring the allocation of scarce resources toward younger patients, as this is expected to save more years of life (Emanuel et al. 2020).

C15.P14 We presented participants with a version of the ventilator dilemma which pits the utilitarian principle against the 'first come, first served' principle. In this case, participants must decide whether to give the last available ventilator to a 65-year-old patient who arrived first, or a 25-year-old patient who arrived a few moments later. (Participants were told to assume a life-expectancy of 80 years for both patients, if saved by the ventilator.) In the VOI stage, participants were asked how they would want the ventilator to be allocated if they knew they had a 50 per cent chance of being the older patient (with 15 years left to live) and a 50 per cent chance of being the younger patient (with 55 years left to live).

C15.P15 As expected, most participants, when engaged in VOI reasoning, favoured giving the ventilator to the younger patient. In other words, they preferred to have (A) a 50 per cent chance of being a 25-year-old who lives another 55 years and 50 per cent chance of dying at age 65, rather than (B) having a 50 per cent chance of being a 65-year-old who lives another 15 years and a 50 per cent chance of dying at age 25. Most critically, the participants who first worked

through the VOI version of this dilemma were subsequently more likely to favour allocating the ventilator to the younger patient when presented with the original dilemma.

C15.P16 In our second experiment using the ventilator dilemma, we replicated the original result using a larger sample. This enabled us to break down the results by the age of the participant, which turned out to be very illuminating. Among younger participants (ages 18–30), the VOI reasoning exercise had little effect: 66 per cent favoured the younger patient in the VOI condition, while 62 per cent favoured the younger patient in the control condition. This is not so surprising, as younger participants may be expected to favour younger patients, with or without VOI reasoning. For participants ages 31–59, the results were stronger: VOI reasoning pushed utilitarian judgments from 47 per cent to 61 per cent. But for participants over age 60, we observed a dramatic reversal: Without engaging in VOI reasoning, only 33 per cent of older participants favoured saving the younger patient. But when older participants engaged in VOI reasoning, 62 per cent subsequently favoured allocating the ventilator to the younger patient. In other words, VOI reasoning completely eliminated self-serving bias in older participants, nearly doubling the number who favoured saving more years of life. The VOI reasoning exercise made their subsequent moral judgments look like those of people in their 20s.

C15.S3

15.3 The psychology of VOI reasoning

C15.P17 Why does VOI reasoning encourage utilitarian responses to these dilemmas? Our suggestion is that people are having a genuine philosophical insight, in the spirit of Rawls (and Harsanyi—see Section 15.4). We think that the VOI manipulation encourages people to think more impartially and, as a result, changes their judgments. To better understand what we have in mind, we'll need to review our current scientific understanding of what goes on in people's minds/brains when they respond to dilemmas such as these. For this, we'll focus on the *footbridge* dilemma and others like it, since they are the best understood.

C15.P18 According to the dual-process theory, there are two competing forces at work. On the one hand, there is impartial cost–benefit reasoning, which depends on conscious, controlled processing dependent on the fronto-parietal control network (Greene et al. 2004; Shenhav and Greene 2014; Conway and Gawronski 2013; Conway et al. 2018; Patil et al. 2020). The footbridge case, however, also involves a more reactive emotional component. Pushing the man off the footbridge is a prototypically violent action. More specifically, pushing entails causing an innocent person's death in a manner that is active, direct, and intended as a means to an end. These factors interact to make people less likely to approve of the utilitarian option in cases such as this (Cushman et al. 2006; Greene et al. 2009; Feltz and May 2017; Patil 2015). As noted above, the mechanism is emotional. This is seen most clearly in the increased utilitarian judgments of patients with emotional deficits (Mendez et al. 2005; Koenigs et al. 2007; 2012; Ciarmelli et al. 2007; Moretto et al. 2012; Thomas, Croft and Tranel 2011; Koven 2011; Patil and Silani 2014) and reduced utilitarian judgments among patients (McCormick et al. 2016), people under the influence of psychoactive drugs (Crockett et al. 2010), and ordinary people (Cushman et al. 2012; Conway and Gawronski 2013; Conway et al. 2018; Gleichgerrcht and Young 2013; Costa et al. 2014; Geipel et al. 2015) with increased reliance on emotional response.

C15.P19 This dual-process dynamic is perhaps best understood, at a computational level, as reflecting the distinction between 'model-free' and 'model-based' modes of learning and decision-making (Sutton and Barto 1998; Daw and Doya 2006), here applied to the domain of moral judgment (Cushman 2013; Crockett 2013; Greene 2017; Patil et al. 2020). In short, we recoil at the thought of committing a violent act, such as pushing someone off a footbridge, because we have learned (through our own experience or vicariously through others) that such actions typically lead to bad outcomes (directly for others and indirectly for ourselves). But in the moment, it's not the expectation of a bad outcome that triggers that response. The negative emotional response is attached to the 'act itself' (Bennett 1995), independent of its current expected consequences, but very much due to the consequences that actions such as this have had in the past. This explains why people are reluctant to perform pretend acts of violence in the lab, even when they are fully aware that no bad consequences will follow (Cushman et al. 2012), and why a rat trained to press a lever for food will continue to do so even when it has entered the cage fully fed (Cushman 2013). The utilitarian response, by contrast, appears to be model-based (Patil et al. 2020). That is, it is based on a causal model of the world—an explicit understanding of which actions will lead to which consequences—and a preference for one set of consequences (five people alive, one dead) over the opposite.

C15.P20 With all of this mind, let's consider how the VOI exercise exerts its influence. Consider the VOI footbridge case: If you don't know who you're going to be (the pushable person on the footbridge or one of the five to be saved on the tracks), what do you want the decision-maker to do? As you consider your self-interested choice between option A (giving you a 5 out of 6 chance of living) and option B (giving you a 1 in 6 chance of living), which factors inform your decision? The data suggest that, when it's your own life at stake, you don't care much about whether death under option A involves pushing, while death under option B does not. Nor do you care about whether you would be killed as a means to an end (option A) or merely as side-effect (option B). Nor do you care about whether these events might appear, to a judgmental onlooker, like a ghastly murder (option A), as opposed to a tragic accident (option B). Nor do you care about what choosing option A over option B might say about the moral character of the decision-maker. Nor do you care about whether option A or option B would be required by the set of rules that would overall make things go best if everyone were to follow them. From a purely self-interested perspective—as required by the VOI procedure—all you really care about is your odds of surviving. The VOI procedure takes the focus off all the subtle psychological factors behind all of the subtle philosophical theories and puts the focus squarely on the consequences.

C15.P21 So, you've decided that, from behind the veil, you want the decision-maker to push because you prefer probably living to probably dying. And now you face the original moral question: is pushing morally acceptable? Even without the VOI experience behind you, there's a clear argument in favour of pushing: Better to save more lives. For some people, that's enough. But for most people, the negative feeling attached to the action carries more weight. This is not an unhealthy sign, since such feelings are responsible for making us behave non-psychopathically (Koenigs et al. 2012; Greene 2013). Within the general population there is a correlation between (self-reported) antisocial tendencies and a willingness to endorse such utilitarian sacrifices (Bartels and Pizarro 2011; Kahane et al. 2015),[5] and people

[5] Bartels and Pizarro (2011) and Kahane et al. (2015) present these findings as a challenge to the dual-process theory, even though they are explicitly predicted by the dual-process theory, consistent with

seem to know intuitively that individuals who endorse such sacrifices are to be viewed with suspicion (Everett et al. 2018). And yet, trusting that intuition means that five people, rather than one, are dead—at least, hypothetically. The VOI gives those five people a voice by putting you (probabilistically) in their shoes. Justifying violence by appeal to the greater good smacks of moral callousness. Life is full of opportunities to justify morally questionable behaviour in this way, which is why 'The ends don't justify the means' is a nugget of folk moral wisdom. But the VOI furnishes a less suspicious justification, unsullied by widespread abuse. When a decision-maker is dealing with a moral dilemma where the utilitarian response is unpalatable, and therefore unpopular, employing a VOI justification of the utilitarian response is more appealing. Compared to a utilitarian justification of the same response, VOI justifications increase observer trust of the decision-maker, an effect driven by perceived warmth (Huang 2020). For example, in response to the footbridge case, one can justify pushing, not as the ends justifying the means, but by appeal to *impartiality*. One can assure oneself—and others, if necessary—that one is not the sort of ruffian who thinks nothing of pushing innocent people to their death. Instead, one can say, honestly and earnestly: *This is what I would want for myself if I did not know who I was going to be.*

C15.S4

15.4 Normative implications

C15.P22 What implications, if any, do these findings have for normative ethics? We've provided evidence that going through the VOI exercise tends to make people's judgments more utilitarian—if not in all cases, then across a substantial range, from self-driving cars to charitable donations. But is this *an improvement*? There are reasons to think that it is.

C15.P23 First, there is a long tradition in moral philosophy according to which judgments are expected to improve with informed reflection (Smith 1994; Smith 1759/2010). And, critically, this enthusiasm for reflection extends far beyond the utilitarian/consequentialist tradition, including, as one of its chief proponents, Rawls (1971), who canonized the method of 'reflective equilibrium'. In most psychological studies of decision-making, the punchline is that human rationality is sorely lacking (Ariely 2008; Kahneman 2003), but here—refreshingly, perhaps—we observe humans gravitating toward a normative ideal, and with no special training. Human judgments are subject to framing effects (Tversky and Kahneman 1981), priming effects (Payne, Brown-Iannuzzi, and Loersch, 2016), arbitrary reference points (Tversky and Kahneman 1974), the influence of incidental emotions (Lerner et al. 2015), and so on. Such influences operate unconsciously, exploiting our biases. But the VOI manipulation isn't 'manipulative'. It's a conscious reasoning exercise aimed at removing bias. It's *Socratic*. It doesn't tell you what to believe or value. Nor does it fly beneath the radar of

AQ: Payne et al. 2016 not in Refs.

prior work (Glenn et al. 2009; Koenigs et al. 2007; 2012; Ciaramelli et al. 2007). Kahane et al. (2015) make the stronger claim—which is genuinely at odds with the dual-process theory—that ordinary people's sacrificial utilitarian judgments are driven *entirely* by antisocial tendencies, i.e. reduced concern about causing harm. However, Conway et al. (2018) re-ran all of Kahane et al.'s (2015) experiments with the addition of process dissociation measures, and showed that ordinary people's sacrificial utilitarian judgments reflect a mixture of antisocial tendencies and genuine concern for the greater good—a combination precisely predicted by the dual-process theory.

reason. It simply asks a question, leaving it to you to formulate your answer and assess its relevance. To the extent that we regard rational reflection as a good influence, we should welcome the effects of VOI reasoning.

C15.P24 Second, these results raise further doubts about the reliability of our anti-utilitarian moral intuitions. According to the conventional wisdom among ethicists, our feeling that it's wrong to push in the footbridge case reflects a genuine philosophical insight. Philosophers such as Elizabeth Anscombe (1958), Bernard Williams (1973/2012), John Rawls (1971), Judith Thomson (1985), Frances Kamm (1998), and Michael Sandel (2010) point to sacrificial dilemmas such as the footbridge case as evidence that there is something wrong with utilitarianism/consequentialism. These anti-utilitarian intuitions, they say, reflect a proper appreciation of countervailing moral concerns, most often framed in terms of individual rights (Thomson 1990). The more sceptical alternative, favoured by Greene (2007; 2013; 2014) and others (Baron 1994; Singer 2005; Sunstein 2005) is that our negative responses to utilitarian sacrifices are overgeneralizations of otherwise good heuristics, encoded in our emotional dispositions. Again, we recoil at acts of violence (and other less dramatically harmful actions) because it is generally good to do so—directly good for others and indirectly good for ourselves. But when philosophers devise devilish dilemmas in which canonically bad actions are guaranteed to produce the best possible results, our emotional dispositions can't adjust (Greene 2017).

C15.P25 On the more optimistic view of anti-utilitarian intuition, one might expect the VOI manipulation to have no effect. If the VOI helps us think more clearly about the demands of justice, and our intuitions are already attuned to the demands of justice, then one might expect the VOI manipulation to be redundant (or, perhaps, to push us even further from the utilitarian response). In other words, you might think that the value of justice is already 'priced in' to our intuitions, leaving nothing for the justice-boosting VOI reasoning to do. But, instead, it leads people to favour the greater good. Why?

C15.P26 From a utilitarian/consequentialist perspective, the answer is straightforward. Impartiality is a central feature of utilitarian/consequentialist thought, and what it means to be impartial is to count everyone's well-being equally in one's assessment of the greater good. To value the life of the person on the footbridge over the lives of five others is, from this perspective, not at all impartial. But there are, of course, non-utilitarian conceptions of impartiality, focused not on maximizing the sum of individual well-being (with each individual counting equally) but on a respect for rights that all people have, including the right not to be used as a trolley-stopper.

C15.P27 Why, then, doesn't the VOI exercise boost this alternative conception of impartiality? The answer, we think, is that our anti-utilitarian intuitions are, at best, only loosely related to impartiality. We suggest that they are not about valuing all people's well-being equally. Rather, they are heuristics for avoiding bad actions. This is not completely unrelated to impartiality because restraints on harmful behaviour tend, overall, to make all of us better off. But the foregoing evidence suggests that our anti-utilitarian intuitions are not about balancing *present* moral considerations in a fair and just way. Instead, they are a low-bandwidth signal about what has been bad in the past.

C15.P28 In assessing the psychology behind the VOI and its normative implications, there is an interesting comparison to a different pro-utilitarian shift. Kurzban, DeScioli, and Fein (2012) presented participants with trolley cases in which the individuals whose lives are at stake are all siblings. Would you kill one of your brothers to save five of your brothers? They found that

people tended to be more utilitarian when the dilemma was all in the family (47 per cent vs 28 per cent). Why? We suggest that this effect, like the VOI effect, comes from shifting the focus onto consequences. Millions of strangers die every day, and we carry on just fine. What's more, most of us (readers of chapters such as this) are in a position to prevent strangers from dying through effective charitable giving (Singer 2010; 2015; MacAskilll 2015), and yet nearly all of us do either nothing or far less than we could. Sad but true, the deaths of strangers, in and of itself, bothers us very little. But actively killing a stranger is very different. When we think about pushing an innocent person to his death, our amygdalae catch fire. Thus, in the footbridge case, where the choice is between passively allowing the deaths of five strangers versus actively killing one, the latter is far more salient. But the deaths of siblings, unlike the deaths of strangers, really matter to us. Each one matters individually, and because each one matters individually, the losses *add up*.

C15.P29 Note the similarity between this utilitarian shift and the one induced by the VOI exercise. Once again, when it's *your own* life at stake, you don't care about whether you might get pushed before you get killed. You just care about being killed. This tendency to get more utilitarian as the personal stakes go up suggests something that many ethicists might find surprising: When you *really care*, you focus on the consequences, and the numbers matter.

C15.P30 A final example: Xin Xiang went to Tibet and interviewed 48 Tibetan Buddhist monks. She presented each of them with a version of the footbridge dilemmas and found that a whopping 83 per cent of them approved of pushing (Xiang 2014; Xiang and Greene 2019). Many of the monks cited a specific sutra about a ship captain. The captain killed a man who was planning to kill many others, expecting that he would suffer a great loss to his karma for performing this terrible act. But because he did it for the sake of others, not for himself, he received divine reward rather than punishment. The monks Xiang interviewed understood the footbridge case as a dilemma, noting that it is, generally speaking, a terrible sin to kill another human. But, they explained, if one does so with the noble intention of helping others, then it is acceptable, even praiseworthy. When we describe these results to people, they are often surprised. They see the deontological response as the moral high road and the utilitarian response as merely 'pragmatic'. And they expect more high-road than pragmatism from high-altitude monks. Those familiar with the trolleyological literature are surprised to find that Buddhist monks respond to the footbridge case with an answer disproportionately favoured by psychopaths (Koenigs et al. 2012) and patients with VMPFC damage (Ciarmelli et al. 2007; Koenigs et al. 2007). But, as you might expect, the monks show no sign of being antisocial or otherwise emotionally compromised. It seems, instead, that their scholarly traditions and meditative practices have led them to value the intention to do the most good, even when it's emotionally uncomfortable. Of course, Buddhist monks are not the ultimate arbiters of right and wrong. But their responses, at the very least, indicate that not all kinds of moral concern manifest in the same way (Conway et al. 2018).

AQ: Xiang 2014 is not in Refs.

C15.P31 What these studies suggest is that consequentialist moral concern may be the deepest kind of moral concern. When we say that it's wrong to push the man off the footbridge, even at a net cost of four lives, we imagine ourselves atop the moral high ground. And compared to antisocial people who might give the same answer, that may be true. But when we do this we are, in a very real sense, ignoring the golden rule. We are not caring about others the way that we care about ourselves. Nor are we treating strangers like brothers or sisters. For ourselves and our loved ones, it's the consequences that matter. Why should we not extend *that* kind of moral concern to everyone (Hare 2016)?

C15.P32 Before moving on, we wish to be clear about an issue that we are *not* addressing here, namely the debate between Rawls and Harsanyi (1955; 1975) over whether VOI reasoning favours a social order based on a utilitarian principle vs Rawls's 'maximin' principle, or some variant thereof. We've provided evidence that veil-of-ignorance reasoning leads people to more utilitarian judgments. This is somewhat surprising because the most celebrated use of veil-of-ignorance reasoning, that of Rawls in *A Theory of Justice*, is the centrepiece of an argument *against* utilitarianism. Rawls argued that citizens, deliberating from behind a veil of ignorance, would reject utilitarian principles as too risky. According to Rawls, in a utilitarian society one could end up oppressed, perhaps even enslaved. Why? Because, according to utilitarianism, doing anything to anyone can be justified as long as it produces enough utility somewhere else. Rawls, instead, favours the 'maximin' principle, which rank orders outcomes based on the well-being of the least well-off person within each outcome.

C15.P33 The economist John Harsanyi, recipient of the Nobel prize for his pioneering work in the field of game theory, devised his own version of the veil-of-ignorance argument, independently of Rawls and at about the same time (Harsanyi 1955). Harsanyi concluded that veil-of-ignorance reasoning favoured utilitarian social principles, as citizens choosing from behind the veil would seek to maximize their respective expected utilities. Harsanyi argued, moreover, that utilitarianism would never, in fact, endorse slavery or other forms of oppression, but that the maximin principle could lead to absurd policies whereby massive gains to millions of people are foregone in order to provision barely perceptible benefits to a much smaller number—perhaps just one person (Harsanyi 1975).

C15.P34 Elsewhere, one of us has argued in favour of Haranyi's view (Greene 2013: 383–5); but, for present purposes, we wish to be clear that the results described above, both our own and those of others, simply do not address the Rawls/Harsanyi debate. This is because the dilemmas we've examined do not clearly separate the utilitarian and maximin principles. For example, in the footbridge case, the utilitarian answer is clear, but what does maximin say? One could argue that being pushed and run over by a trolley is worse than simply being run over by a trolley, but this is a tenuous assumption at best, and it could easily be eliminated with minor tweaks to the dilemma (e.g. pushing the person into the trolley's path down a slide, yielding an enjoyable ride). It's not clear what the maximin principle says about the footbridge case or any of the cases that we've discussed. Thus, at present, we make no claims about whether Rawls is right that VOI reasoning favours maximin over a utilitarian principle. (But see Frohlich, Oppenheimer, and Eavey 1987 for experimental results that address this question.) Instead, we claim only that veil-of-ignorance reasoning favours the greater good across a range of dilemmas, including some with real-world significance.

C15.S5 15.5 A TOOL FOR THINKING ABOUT REAL-WORLD PROBLEMS

C15.P35 Once again, our interest in veil-of-ignorance reasoning is not to defend pushing hypothetical people off hypothetical footbridges, but instead to develop a useful tool for thinking about real-world moral problems. Real-world moral problems involve difficult trade-offs, and sometimes the policy that is expected to do the most good, or to be the most fair, is not

the one that feels the most right. VOI reasoning, we propose, can help us distinguish the things that really matter from the things that exploit our intuitive biases. Put in Rawlsian terms, we think that VOI reasoning, applied to specific moral dilemmas, can help us find our reflective equilibrium (Greene 2014).

C15.P36 VOI reasoning may be useful from two perspectives. First, for those of us who are already committed to promoting the greater good, VOI reasoning is useful insofar as it encourages others to do the same. Second, for those of us who are committed to impartiality, but not necessarily to promoting the greater good, VOI reasoning is useful insofar as it helps us think more impartially, wherever that may lead.

C15.P37 Let's return to the case of utilitarian AVs, swerving for the greater good. Do the moral challenges posed by AVs in the real world have anything to do with the moral dilemmas considered here? It may be tempting to dismiss such stylized dilemmas as irrelevant (e.g. Roberts 2018). First, AVs will rarely, if ever, face such stark choices between, say, killing one person and killing five others. AVs will soon be, if they are not already, far more perceptive and deft drivers than humans. As a result, they will avoid such situations before they arise. What's more, they won't use simple rules of the kind framed by philosophers or lawyers. Instead, they will use complex machine learning algorithms, applying incomprehensibly subtle, context-sensitive dispositions that have been acquired through millions of hours of driving experience. And on those rare occasions when AVs find themselves in trolley-like moral dilemma, whatever mistakes they might make will be minimal compared to the thousands of lives they save each year by being generally better drivers than us. (In the US alone, human drivers kill over 35,000 people per year: US DOT 2018.) Thus, fretting over the AV trolley problem, if it delays the arrival of superior driving machines by one day, is itself a bigger problem than the AV trolley problem will ever be.

C15.P38 There is much truth in these assertions, but not enough to make the AV trolley problem go away. First, while it's true that AVs will rarely face stark choices between killing one and killing five, the same underlying dilemma re-emerges as a set of questions about how to apportion risk. Such decisions are familiar to human drivers: You're driving behind a cyclist on a narrow two-lane highway. There's steady stream of oncoming traffic. You could *probably* zip around the cyclist before the next car gets too close. Do you go for it? Or do you wait (and wait? … and wait?) for a wider window to open? It's true that there is no clear moral principle that can tell you when it's morally acceptable, or not, to zip around. But it doesn't follow from this that there's no moral decision to be made. In deciding when, whether, and how you are willing to pass a vulnerable cyclist, you are making a decision about how much you are willing to risk harming, possibly killing, another human. What's more, driving routinely involves these sorts of probabilistic micro-dilemmas. Human drivers can't avoid the question: Will I drive nicely or like a jerk? Why, then, should we think that driving machines, and the people who design them, can avoid this question? It's true that machines will be more adept drivers than humans, but the humans still have to decide what counts as more morally adept. If a self-driving car, in the course of training its navigational neural network, never hits a cyclist, but misses a cyclist by less than six inches in 3 per cent of cases, is that a policy to be reinforced or revised? That's not a question that the car, or its superhumanly subtle navigation system, can answer. Humans must supply the moral standard.

C15.P39 As an unwitting Mercedes Benz executive discovered, there is no intuitively appealing moral standard for AVs (Morris 2016). Christoph von Hugo was asked whether Mercedes Benz's self-driving cars would prioritize the safety of their passengers over others.

Hugo presented his privilege-the-passengers position as a matter of consequentialist commonsense: 'If you know you can save at least one person, at least save that one. Save the one in the car.' But there's no reason to assume that 'the one you can save' will always be 'the one in the car', as, for example, when vulnerable pedestrians and cyclists are involved. Von Hugo's comments caused a minor uproar. Mercedes quickly realized that this was a no-win situation. Do we privilege the lives of our already privileged passengers? Or do we say that we're willing to sacrifice our passengers for the greater good of others? The automaker eventually issued a statement: 'Neither programmers nor automated systems are entitled to weigh the value of human lives' (Daimler 2018). But that position is simply untenable, as anyone who has ever waited patiently behind a cyclist knows. Not weighing is not an option.

C15.P40 Thus, difficult moral choices will be made, even if they are less stark and more probabilistic than classic trolley dilemmas. Fortunately, the VOI argument applies to these more graded dilemmas as well. In the VOI dilemmas that we used in our experiments, one chooses between outcomes in which one has high vs low odds of surviving. Making the original dilemma probabilistic makes the VOI dilemma a choice between having a high probability of having a high probability of surviving vs having a low probability of having a low probability of surviving. The mathematics is more complicated, but the upshot will likely be the same. More generally, when thinking about these problems, we can still ask the question: what would you want if you didn't know who you were going to be? AV algorithms that minimize the loss of life, even probabilistically, will probably pass this test.

C15.P41 As noted earlier, approval of the utilitarian AV policy (requiring cars to value all lives equally) rose to over 80 per cent following VOI reasoning. This finding is notable because it makes significant progress toward resolving what Bonnefon et al. (2016) call the 'Social Dilemma of Autonomous Vehicles', whereby people espouse general approval of utilitarian AVs that value all lives equally, but disapprove of policies that would require AVs to value all lives equally (and not privilege passengers over others). The VOI seems to move people strongly toward approval of such policies. Although this finding does not speak directly to the question of whether people would choose to *ride* in such vehicles, one might hope that if such a policy were approved and enacted, ridership would follow.

C15.P42 Next consider the case of charitable giving. The effect of VOI reasoning on effective giving is interesting, in part, because one might not expect it to work. Recall that our participants were asked to decide between an American charity expected to restore vision in one person's eye and an Indian charity expect to restore vision in two people's eyes. As expected, most people in the VOI condition said that they would prefer the money go to the more effective charity, giving them a 2 in 3 chance of getting treatment, rather than a 1 in 3 chance. But why, you might wonder, should this VOI judgment transfer to the actual donation decision? After all, one could say, 'Yes, I would prefer better odds for myself, but this decision about where to donate is not about helping me. I feel a greater obligation to my fellow Americans than I do to people in India.' In other words, one might not expect the VOI judgment to transfer to the real donation decision because the real donation decision, unlike the VOI decision, involves a choice between in-group and out-group. And yet a significant subset of our participants were moved to be less tribalistic by thinking about this decision from a more impartial perspective.

C15.P43 Does this matter? Americans alone give about $400 billion to charity each year (Giving USA 2018). If even 1 per cent of that amount were directed toward more effective charities, that would be about $4 billion per year. According to GiveWell, saving a life today probably costs about $3,000, if the money is spent well (GiveWell 2016). Thus, a 1 per cent shift toward

truly effective giving could save over 1,000 lives—very possibly fewer, but very possibly more. Thus, small shifts in how people think about charitable giving may be important, and it seems that VOI reasoning can produce small shifts. We are not claiming, of course, that we have produced a scalable method for directing money to effective charities. Our point is simply that in the domain of charitable giving, the stakes are high enough and wide enough that small changes in how people think can be a matter of life and death.

C15.P44 Our experiments using the ventilator dilemma are the most directly applicable of all, as this dilemma was all too real for Italian doctors in 2020 (Mounk 2020) and is likely to reappear in other places. (At the time of writing, COVID-19 is spreading rapidly through nations such as Brazil, where medical resources are often scarce.) What's more, there is genuine disagreement about whether age should be a factor in such cases. Roger Severino, the director of the Office for Civil Rights at the U.S. Department of Health and Human Services, described policies that allocate resources based on age as 'ruthless utilitarianism' and announced his intention to investigate those who apply them (Fink 2020). But, as we've shown, older people—the people whose rights Severino hopes to protect—become a lot more 'ruthless' when they consider this problem from behind a veil of ignorance. Once again, what seems like callous ageism may instead reflect the truest application of the golden rule—caring about others the way one cares about oneself, while giving equal weight to everyone.

C15.P45 We've focused on AVs, charitable giving, and healthcare because they are featured in our experimental work, but there are surely other real-world moral dilemmas that are amenable to change through VOI reasoning. Would people favour a hefty wealth tax, more inclusive immigration reform, stricter carbon emissions standards, or more expansive rights for minorities if people didn't know who among those affected they would be? Future research awaits. We suspect, however, that the greater challenge lies not in demonstrating such proof-of-principle effects in controlled experiments, but in figuring out how to put such effects to work in the real world.

C15.S6 References

C15.P46 Anscombe, G. E. M. 1958. Modern moral philosophy. *Philosophy* 33: 1–19.

C15.P47 Ariely, D. 2008. *Predictably Irrational*. New York: Harper.

C15.P48 Baron, J. 1994. Nonconsequentialist decisions. *Behavioral and Brain Sciences* 17(1): 1–10.

C15.P49 Bartels, D. M., and D. A. Pizarro. 2011. The mismeasure of morals: antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition* 121(1): 154–61.

C15.P50 Bennett, J. (1995). *The Act Itself*. Oxford: Clarendon Press.

C15.P51 Bonnefon, J. F., A. Shariff, and I. Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* 352(6293): 1573–6.

C15.P52 Ciaramelli, E., M. Muccioli, E. Làdavas, and G. di Pellegrino. 2007. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience* 2(2): 84–92.

C15.P53 Conway, P., and B. Gawronski. 2013. Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. *Journal of Personality and Social Psychology* 104(2): 216.

C15.P54 Conway, P., J. Goldstein-Greenwood, D. Polacek, and J. D. Greene. 2018. Sacrificial utilitarian judgments do reflect concern for the greater good: clarification via process dissociation and the judgments of philosophers. *Cognition* 179: 241–65.

C15.P55 Costa, A., A. Foucart, S. Hayakawa, M. Aparici, J. Apesteguia, J. Heafner, and B. Keysar. 2014. Your morals depend on language. *PLoS ONE* 9(4): e94842.

C15.P56 Crockett, M. J. 2013. Models of morality. *Trends in Cognitive Sciences* 17(8): 363–6.

C15.P57 Crockett, M. J., L. Clark, M. D. Hauser, and T. W. Robbins2010. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences* 107(40): 17433–8.

C15.P58 Cushman, F. 2013. Action, outcome, and value: a dual-system framework for morality. *Personality and Social Psychology Review* 17(3): 273–92.

C15.P59 Cushman, F., K. Gray, A. Gaffey, and W. B. Mendes. 2012. Simulating murder: the aversion to harmful action. *Emotion* 12(1): 2.

C15.P60 Cushman, F., L. Young, and M. Hauser. 2006. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science* 17(12): 1082–9.

C15.P61 Daimler Global Media. 2018. Daimler clarifies: Neither programmers nor automated systems are entitled to weigh the value of human lives. https://media.daimler.com/marsMediaSite/en/instance/ko/Daimler-clarifies-Neither-programmers-nor-automated-systems-are-entitled-to-weigh-the-value-of-human-lives.xhtml?oid=14131869

C15.P62 Daw, N. D., and K. Doya. 2006. The computational neurobiology of learning and reward. *Current Opinion in Neurobiology* 16(2): 199–204.

C15.P63 Emanuel, E. J., G. Persad, R. Upshur, et al. 2020. Fair allocation of scarce medical resources in the time of Covid-19. *New England Journal of Medicine* 382: 2049–55.

C15.P64 Everett, J. A., N. S. Faber, J. Savulescu, and M. J. Crockett. 2018. The costs of being consequentialist: social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology* 79: 200–216.

C15.P65 Feltz, A., and J. May. 2017. The means/side-effect distinction in moral cognition: a meta-analysis. *Cognition* 166: 314–27.

C15.P66 Fink, S. 2020. U.S. Civil Rights Office rejects rationing medical care based on disability, age. *New York Times*, 30 Mar. https://www.nytimes.com/2020/03/28/us/coronavirus-disabilities-rationing-ventilators-triage.html

C15.P67 Frohlich, N., and J. A. Oppenheimer. 1993. *Choosing Justice: An Experimental Approach to Ethical Theory*. Berkeley: University of California Press.

C15.P68 Frohlich, N., J. A. Oppenheimer, and C. L. Eavey. 1987. Laboratory results on Rawls's distributive justice. *British Journal of Political Science* 17(1): 1–21.

C15.P69 Geipel, J., C. Hadjichristidis, and L. Surian. 2015. How foreign language shapes moral judgment. *Journal of Experimental Social Psychology* 59: 8–17.

C15.P70 GiveWell. 2016. GiveWell cost-effectiveness analysis—November 2016. https://docs.google.com/spreadsheets/d/1KiWfiAGX_QZhRbC9xkzf3I8IqsXC5kkr-nwY_feVlcM/edit#gid=1034883018

C15.P71 Giving USA. 2018. Americans gave $410.02 billion to charity in 2017, crossing the $400 billion mark for the first time. https://givingusa.org/giving-usa-2018-americans-gave-410-02-billion-to-charity-in-2017-crossing-the-400-billion-mark-for-the-first-time/

C15.P72 Gleichgerrcht, E., and L. Young. 2013. Low levels of empathic concern predict utilitarian moral judgment. *PLoS ONE* 8(4): e60418.

C15.P73 Glenn, A. L., A. Raine, and R. A. Schug. 2009. The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry* 14(1): 5–6.

C15.P74 Greene, J. D. 2007. The secret joke of Kant's soul. *Moral Psychology* 3: 35–79.

C15.P75 Greene, J. D. 2013. *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*. London: Penguin.

C15.P76 Greene, J. D. 201). Beyond point-and-shoot morality: why cognitive (neuro) science matters for ethics. *Ethics* 124: 695–726.

C15.P77 Greene, J. D. 2017. The rat-a-gorical imperative: moral intuition and the limits of affective learning. *Cognition* 167: 66–77.

C15.P78 Greene, J. D., F. A. Cushman, L. E. Stewart, K. Lowenberg, L. E. Nystrom, and J. D. Cohen 2009. Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition* 111(3): 364–71.

C15.P79 Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293(5537): 2105–8.

C15.P80 Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44(2): 389–400.

C15.P81 Hare, C. 2016. Should we wish well to all? *Philosophical Review* 125(4): 451–72.

C15.P82 Huang, K. 2020. Third-party judgments of veil-of-ignorance reasoning. In *Veil-of-Ignorance Reasoning and Justification of Moral Judgments*. Doctoral dissertation, Harvard University.

C15.P83 Huang, K., R. Bernhard, N. Barak-Corren, M. Bazerman, and J. D. Greene. 2020. Veil-of-ignorance reasoning favors allocating resources to younger patients during the COVID-19 crisis. MS. DOI: 10.31234/osf.io/npm4v

C15.P84 Huang, K., J. D. Greene, and M. Bazerman. 2019. Veil-of-ignorance reasoning favors the greater good. *Proceedings of the National Academy of Sciences* 116(48): 23989–95.

C15.P85 Galinsky, A. D., and G. B. Moskowitz. 2000. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology* 78(4): 708.

C15.P86 Harsanyi, J. C. 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy* 63(4): 309–321.

C15.P87 Harsanyi, J. C. 1975. Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory. *American Political Science Review* 69(2): 594–606.

C15.P88 Kahane, G., J. A. Everett, B. D. Earp, M. Farias, and J. Savulescu. 2015. 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition* 134: 193–209.

C15.P89 Kahneman, D. 2003. A perspective on judgment and choice: mapping bounded rationality. *American Psychologist* 58(9): 697.

C15.P90 Kamm, F. M. 1998. *Morality, Mortality*, vol. 1: *Death and Whom to Save From It*. New York: Oxford University Press.

C15.P91 Koenigs, M., M. Kruepke, J. Zeier, and J. P. Newman. 2012. Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience* 7(6): 708–14.

C15.P92 Koenigs, M., L. Young, R. Adolphs, et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446(7138): 908.

C15.P93 Koven, N. S. 2011. Specificity of meta-emotion effects on moral decision-making. *Emotion* 11(5): 1255.

C15.P94 Kurzban, R., P. DeScioli, and D. Fein. 2012. Hamilton vs. Kant: pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior* 33(4): 323–33.

C15.P95 Lerner, J. S., Y. Li, P. Valdesolo, and K. S. Kassam (2015). Emotion and decision making. *Annual Review of Psychology* 66: 799–823.

C15.P96 MacAskill, W. 2015. *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*. London: Guardian Books.

C15.P97 McCormick, C., C. R. Rosenthal, T. D. Miller, and E. A. Maguire. 2016. Hippocampal damage increases deontological responses during moral decision making. *Journal of Neuroscience* 36(48): 12157–67.

C15.P98 Mendez, M. F., E. Anderson, and J. S. Shapira (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology* 18(4): 193–7.

C15.P99 Moretto, G., E. Làdavas, F. Mattioli, and G. Di Pellegrino. 2010. A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience* 22(8): 1888–99.

C15.P100 Morris, D. Z. 2016. Mercedes-Benz's self-driving cars would choose passenger lives over bystanders. *Fortune*, 15 Oct.

C15.P101 Mounk, Y. 2020. The extraordinary decisions facing Italian doctors. *The Atlantic*, 11 Mar. https://www.theatlantic.com/ideas/archive/2020/03/who-gets-hospital-bed/607807/

C15.P102 Patil, I., and G. Silani. 2014. Reduced empathic concern leads to utilitarian moral judgments in trait alexithymia. *Frontiers in Psychology* 5: 501.

C15.P103 Patil, I., M. M. Zucchelli, W. Kool, et al. 2020. Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology* 120(2).

C15.P104 Roberts, D. 2018. Don't worry, self-driving cars are likely to be better at ethics than we are. *Vox*, Jan 17.

C15.P105 Singer, P. 2005. Ethics and intuitions. *Journal of Ethics* 9(3-4): 331–52.

C15.P106 Singer, P. 2010. *The Life You Can Save: How to Do Your Part to End World Poverty*. New York: Random House.

C15.P107 Singer, P. 2015. *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. Melbourne: Text Publishing.

C15.P108 Sutton, R. S., and A. G. Barto. 2018. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

C15.P109 Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

C15.P110 Robichaud, C. 2015. Liberty hospital simulation. Classroom exercise.

C15.P111 Sandel, M. J. 2010. *Justice: What's the Right Thing to Do?* London: Macmillan.

C15.P112 Shenhav, A., and J. D. Greene. 2014. Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience* 34(13): 4741–9.

C15.P113 Smith, A. 1759/2010. *The Theory of Moral Sentiments*. London: Penguin.

C15.P114 Smith, M. R. 1994. *The Moral Problem*. Oxford: Wiley-Blackwell.

C15.P115 Sunstein, C. R. 2005. Moral heuristics. *Behavioral and Brain Sciences* 28(4): 531–41.

C15.P116 Thomas, B. C., K. E. Croft, and D. Tranel. 2011. Harming kin to save strangers: further evidence for abnormally utilitarian moral judgments after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience* 23(9): 2186–96.

C15.P117 Thomson, J. J. 1985. The trolley problem. *Yale Law Journal* 94: 1395.

C15.P118 Thomson, J. J. 1990. *The Realm of Rights*. Cambridge, MA: Harvard University Press.

C15.P119 Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185(4157): 1124–31.

C15.P120 Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211(4481): 453–8.

C15.P121 U.S. Department of Transportation. 2018. Traffic safety facts: research note. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812603

C15.P122 Williams, B. 1973/2012. A critique of utilitarianism. In *Ethics: Essential Readings in Moral Theory*, ed. G. Sher. Abingdon: Routledge.

C15.P123 Xiang, X., and J. D. Greene. 2019. Would the Buddha push the man off the footbridge? Exceptionally high levels of utilitarian judgment among Tibetan Buddhist monks. MS.

AQ: Please add a web address.

AQ: Can the reader acccess this MS anywhere?