2

Intuitive and Counterintuitive Morality

Guy Kahane

1. Intuition and Deliberation in Moral Psychology

Recent empirical work on moral psychology is often claimed to have yielded surprising and disturbing results. It has discovered, it is said, that moral judgment is not based in reason or conscious reflection, but in immediate intuitions. And it tells us that individuals rarely engage in deliberation when they make moral judgment, and even when they do, their reasoning merely aims to rationalize a pre-established intuitive conclusion. To shift to psychological terms, the emerging picture is that moral judgment is based almost exclusively in processes that are *automatic*—fast, effortless, and unconscious—and only rarely, if at all, in processes that are *deliberative*—slow, effortful, and involving explicit conscious thinking (Haidt 2001, 2007, 2012; Greene and Haidt 2002).[1]

When these findings are discussed, it's usually assumed that what is surprising about them is the key role they ascribe to intuition in moral judgment. But in itself this is neither surprising nor disturbing. It is hardly news to moral philosophers that intuition plays a central role in our moral lives. Moral philosophers would be surprised, however, to find out that our intuitions are not just starting points for ethical reflection, but also where it invariably ends. A pressing question, then, is whether moral judgment is ever also based in genuine deliberation.

An important body of research by Joshua Greene and his colleagues, and research influenced by it, has offered a striking answer to this question. While broadly confirming the view that automatic processing drives the moral judgments of most individuals, it also appears to show that there is a minority whose moral judgments *are* based on deliberation. Moreover, this minority uses deliberation to arrive at moral conclusions that, far from being mere rationalizations of prior intuitions, actually go *counter*

[1] For general discussion of these two kinds of psychological processes, see Stanovich 1999; Evans 2007.

to their immediate intuitions (Greene et al. 2001; Greene et al. 2004; Greene 2008; Paxton and Greene 2010).

What is most striking about this research, however, is that it has tied the psychological divide between automaticity and "controlled" deliberation to moral judgments with opposing *contents*. For it appears to show that when the majority follow their immediate intuitions, the result is deontological in content, whereas when individuals do engage deliberative reasoning, they arrive instead at contrary utilitarian conclusions.[2] This dual process model of moral judgment is meant as a quite general account of moral psychology, dividing the mind into two modes of processing that generate conflicting moral outputs—and moreover moral outputs that have a strong philosophical resonance.[3]

Needless to say, Greene's dual model would be of great interest even if it were only a descriptive account of our moral psychology. But much more might be at stake here, since Greene and others have gone on to argue that this theory has dramatic normative implications—that it offers support for utilitarianism and its many counterintuitive implications (Greene 2008; Singer 2005).[4]

Notice again that what is potentially disturbing about this model for non-utilitarians *isn't* the role it ascribes to intuition in deontology. Most (though not all) non-utilitarian views explicitly appeal to intuitions. This is merely a truism, something we don't need fancy neuroimaging to know.[5] What might be disturbing isn't the presence of intuition, but the supposed absence of deliberation: the claim that deontological judgments

[2] In what follows, I will reluctantly follow Greene and pretty much everyone else in this literature in using "utilitarian judgment" to refer to what are really just judgments that, in a given decision context, are more utility maximizing than the alternatives (see Kahane and Shackel 2010). As we shall soon see, this terminology is highly misleading at best. I will also reluctantly follow this literature in using "utilitarianism" to refer to what is at best a very simple form of Act Utilitarianism, understood, moreover, not just as a criterion of rightness but also as a concrete decision procedure (see again Kahane and Shackel 2010). This simple variant of utilitarianism is actually rejected by many contemporary utilitarians, let alone by non-utilitarian consequentialists. But the criticism I will be developing here is meant to apply even if we assume this incredibly narrow understanding of utilitarianism. Finally, again in line with the current literature, I will be using "deontological" to refer merely to views that are non-utilitarian.

[3] Greene and others often refer to the view as the "dual process model of moral judgment." But this label is misleading. Greene isn't merely applying the general dual process approach to the mind to the moral domain—something that can be done in multiple ways. What is interesting (and controversial) about Greene's model is the way he ties automaticity and deliberation (aka System 1 and System 2) to distinct moral outputs, something that is in no way entailed by the general dual process approach (see Kahane 2012 for an attempt to precisely state Greene's model and its different components). When I criticize Greene's dual process model in what follows, I intend to refer only to this more specific view; I will later actually outline an alternative way to apply the dual process approach to the model domain.

[4] For criticism of Singer and Greene's normative arguments, see Berker 2009; Kahane, 2010.

[5] It would be slightly, but only slightly, more surprising to discover that non-utilitarian judgments are driven by emotion. But this is something that some non-utilitarians (think of Bernard Williams) explicitly avow. And the evidence suggesting that non-utilitarian judgments are *driven* (as opposed to influenced) by emotion is far from conclusive (see, e.g., Huebner et al. 2009). In any event, whether or not such judgments involve emotion, we now know they are responsive to the interaction of a complex range of factors—they are nothing like crude gut reactions. In line with that, in a recent study we found that a pharmacological intervention that reduces the physiological arousal component of emotion (that is, literally the "gut" part of emotion) in fact led to an *increase* in deontological judgment (Terbeck et al. 2013).

involve nothing *beyond* intuition, and that deliberation always (or perhaps nearly always) points in a utilitarian direction.

In what follows, I will therefore largely grant the claim that deontological judgment is based in intuition.[6] What I will challenge in this chapter is Greene's more controversial claim that utilitarian judgments are uniquely based in deliberative processing. I will argue that a closer inspection of the evidence supports a very different picture of our moral psychology.

My argument will proceed as follows. In section 2, I will provide evidence that there is in fact nothing distinctively utilitarian about deliberation to "utilitarian" conclusions. In section 3, I will go further and argue that such deliberation to "utilitarian" conclusions is in fact best understood as *non*-utilitarian in character. In section 4, I will argue that when individuals do arrive at counterintuitive conclusions in a way that might actually resemble utilitarian reasoning, these judgments actually turn out to be *non-deliberative*—and *worse*. In section 5, I will present findings that suggest that there is actually little connection between so-called "utilitarian" responses to trolley-like dilemmas and genuine concern for the greater good. I conclude, in section 6, by drawing some general lessons for the cognitive science of ethics.

I will not directly consider the possible normative significance of this body of research. But if my argument is even partly successful, then the idea that this empirical research can be used to support utilitarianism will seem highly misguided, at best.

2. Utilitarian or Counterintuitive?

2.1. The Evidence Linking Utilitarian Judgment and Deliberative Processing

There is a great deal of empirical research that is supposed to support Greene's dual process model. However, much of this research in fact only provides evidence for the uncontroversial tie between deontological judgment and automaticity, or for the slightly more controversial tie between such judgment and emotion (see e.g. Greene et al. 2001; Greene et al. 2004; Mendez et al. 2005; Koenigs et al. 2007; Ciaramelli et al. 2007; Moretto et al. 2009).

The available evidence for deliberative processing in utilitarian judgment is rather more limited. Some of it comes from Greene's neuroimaging studies. These studies have reported that moral judgments in so-called "impersonal" trolley-like dilemmas—dilemmas to which most people give "utilitarian" answers—recruit greater activity in areas classically associated with cognitive processing (the right dorsolateral prefrontal cortex (DLPFC) and inferior parietal lobe) compared to "personal" dilemmas like the famous Footbridge case, to which most people give "deontological answers" (Greene

[6] For further discussion of the relation between deontological judgement, intuition, and emotion, see Kahane 2012.

et al. 2001; Greene et al. 2004).[7] Moreover, when subjects do endorse utilitarian solutions to "difficult" personal dilemmas, these judgments turn out to be associated with greater activation in the DLPFC and the dorsal anterior cingulate cortex (dACC), an area implicated in conflict, compared to contrary deontological judgments (Greene et al. 2004).

The classic marker of the automatic/deliberative contrast is difference in response times (RT)—deliberative processing should take longer. Although Greene originally reported that utilitarian judgments in personal dilemmas were associated with greater RTs compared to deontological ones (Greene et al. 2001; Greene et al. 2004), these results have not held up when the stimuli used were better (Moore et al. 2008; Greene et al. 2008; see also McGuire et al. 2009; Greene 2009). However, another classical marker is interference through cognitive load: A competing task should affect only responses based in deliberative processing. Greene et al. 2008 report that a cognitive load manipulation raised response times for utilitarian judgments but not for deontological judgments in "high conflict" personal dilemmas. However, contrary to what Greene's model predicts, such a manipulation did not affect the rates of utilitarian judgment. Nevertheless, in line with the model, more recent studies did find that time pressure increased rates of utilitarian judgment (Suter and Hertwig 2011), whereas stress reduced them (Youssef et al. 2012).

A further source of evidence comes from studies tying utilitarian judgement to traits and capacities associated with deliberative processing. In particular, higher rates of utilitarian judgments were associated both with "need for cognition," a motivational tendency to prefer effortful cognition (Bartels 2008), and with greater working memory capacity, a component of general intelligence (Moore et al. 2008; but see also Koven 2011).[8] In addition, exposure of people to non-moral problems that were specifically designed to assess the capacity to suppress a strong intuitive answer and instead endorse a counterintuitive solution (Frederick 2005) also induced greater rates of utilitarian judgment (Paxton et al. 2012).[9]

[7] In "personal" moral dilemmas like Footbridge, one must choose whether to significantly harm someone in an "up close and personal" way, or to let a larger number of people die; "impersonal" dilemmas involve a similar decision, but the harm is done in a more indirect way. In what follows, I will ignore important worries about classifying moral dilemmas in this way (for further discussion, see Kahane and Shackel 2010).

[8] Notice however that these studies offer only indirect evidence for the role of deliberative processing in utilitarian judgment, since unlike the studies cited here, these studies don't directly measure the processes that are involved in generating such judgments. The tie between utilitarian judgment and, say, a motivational tendency to effortful cognition *may* be explained by greater engagement in effortful cognition when individuals make utilitarian judgments, but it can also be explained by other factors (individuals higher on need for cognition may be different from other people in all sorts of ways that might affect their moral judgment).

[9] As I point out in Kahane 2012, n. 10, Moore et al. 2008 and Paxton et al. 2012 (as well as Paxton et al. 2013) tied "utilitarian" judgment to potential markers of deliberative processing *only* in "personal" dilemmas where—unlike the classic Footbridge case—even if you don't harm someone to save five, that person will die *anyway*. This is problematic, since many deontological views would also endorse such a conclusion. For other methodological worries about some of the evidence listed in this section, see Kahane and Shackel 2011 and Berker 2009.

2.2. *Problematic Inferences*

There is thus considerable evidence appearing to establish a tie between utilitarian judgment and deliberative processing. There is, however, an obvious gap in this evidence. Greene's dual process model is clearly meant to be entirely general: It makes a general claim about utilitarian judgment, a claim that is supposed to apply across moral contexts and domains.[10] But virtually all of the research taken to support the model has focused on so-called "trolley problems" and similar scenarios where one person must be seriously harmed to save a greater number. Such dilemmas of course relate to a rather unusual part of morality, and are merely one of numerous contexts in which utilitarianism clashes with commonsense intuitions[11]—to name just a few other examples, act utilitarianism requires complete impartiality, implying both that we shouldn't give any inherent moral priority to those dear and near to us, and that we have obligations to make extremely demanding sacrifices to help distant strangers in need (I'll return to this example below); utilitarianism also rejects retributive punishment and the notion of desert quite generally, and similarly gives no intrinsic weight to considerations of distributive justice—and this is just the beginning of a long list.[12]

So the model currently relies on a huge inductive leap from findings about trolley-like problems to general conclusions about the sources of "utilitarian" and "deontological" judgment. It's unfortunately common for researchers to report that some study has shown utilitarian judgments to be this or that, where it at most shows this about utilitarian (viz. utility maximizing) judgments *in trolley-like dilemmas*.

We shall later return to this problem. My more immediate aim in what follows, however, is to show that *even* in the narrow domain of trolley-like problems, this supposed tie between utilitarian judgment and deliberative processing involves a serious misreading of the current evidence. In the next two sections, I will therefore focus on

[10] The intended strength of this claim isn't entirely clear. But the idea seems to be that utilitarian judgments are the *favored outputs* of deliberative processing, and that contrary deontological outputs are at best rare exceptions, even aberrations. Notice moreover that Greene's model makes a claim about the *causal source* of each type of judgment. It is perfectly compatible with deontological judgments sometimes *engaging* deliberative processing in an extensive way, so long as this processing plays a merely epiphenomenal role in the final output—so long as it is engaged merely to offer a rationalization of the initial intuition.

[11] In fact, although Greene and other researchers often present the trolley case as a core part of the traditional case *against* utilitarianism, this is highly misleading at best. The trolley problem is a problem *internal* to non-utilitarian ethics, as a look at Foot 1978 and Thomson 1976 would confirm.

[12] When we spell out the ways in which utilitarianism clashes with commonsense morality, we need to be careful with the distinction between criterion of rightness and decision procedure (see n. 2). So-called "indirect" utilitarianism, for example, may allow us to give priority to our loved ones, because such a "decision procedure" (i.e. actual moral psychology) will in fact lead to better consequences. But this point, of course, also applies to how we respond to the classical trolley cases (see Kahane and Shackel 2011). As I explained above, I will be following Greene in using "utilitarianism" to refer to a very simple form of Act Utilitarianism—a slightly simplified version, if you want, of the view associated with Peter Singer.

another problematic inference that underlies the dual process model. Greene and others slide from the supposed empirical finding that

(1) when individuals make utilitarian judgments in trolley cases, deliberative processing plays a causal role in the generation of these judgments

to the conclusion that

(2) deliberative processing reflects the distinctively *utilitarian character* of these judgments.

In what follows, I shall argue that this deliberative processing is unlikely to have much to do with the utilitarian content of these judgments. In fact, we shall see that the connection between deliberative processing and utilitarian judgment is rather superficial.

2.3. An Overlooked Confound?

Here is one obvious reason why it is invalid to infer, from evidence associating "utilitarian" judgments with deliberative processing, the conclusion that this processing reflects the distinctive utilitarian character of these judgments. Greene and other researchers typically compare utilitarian judgments that are highly counterintuitive (such as: *push the fat man* in the Footbridge dilemma) with contrary deontological judgments that are strongly intuitive (e.g. *don't push*). They observe various behavioral and neural differences between the two. And they conclude that these differences must reflect differences between utilitarian and deontological judgments. But there is a glaring alternative explanation they have overlooked: These differences might merely reflect differences between *counterintuitive* and *intuitive* judgments, quite regardless of the content of these judgments. If this is correct, then the apparent tie between process and content is really just an artefact of the kinds of scenarios that researchers have studied, reflecting nothing very interesting about utilitarian and deontological judgments.

Now this alternative explanation wouldn't be a problem if all counterintuitive judgments were utilitarian, and all intuitive ones deontological. But this is obviously not so, as Kant's notorious assertion that we mustn't lie even to prevent murder demonstrates (Kant 1797/1966; this example is actually mentioned in Greene 2008). Most people seem to find it immediately *obvious* that we should lie in such a case—it is natural to say that they find this "utilitarian" decision *intuitive*, and the contrary deontological one strongly *counterintuitive*. And the mere existence of these overlooked pairings of content (utilitarian/deontological) and intuitiveness (intuitive/counterintuitive) already presents apparent counterexamples to Greene's model.

Could Greene deny that there really are these counterexamples? With respect to supposed instances of intuitive utilitarian judgments, he might argue that although some "utilitarian" judgments are immediate and effortless, and based on intuition in some broad sense, they are nevertheless not genuinely automatic—though this would also seem to drain talk of automaticity of much of its content. In any event, Greene himself speculates that utilitarian judgments have their origin in immediate affective

responses to harm to others (Cushman, Young, and Greene 2010). So he already seems committed to accepting this counterexample, although it's not clear how exactly to square this with his general model.[13]

I'm more interested, however, in the possibility of counterintuitive deontological judgments, since they challenge the crucial tie between deliberative processing and utilitarianism. How might Greene challenge this possibility? Greene doesn't deny that some deontological claims, such as Kant's remarks on lying, seem counterintuitive. But he seems to treat these as rare cases where a commitment to an explicit philosophical theory leads some philosophers to make counterintuitive judgments (Greene 2008, 65–6; Paxton and Greene 2010). However, non-philosophers plainly make similar deontological judgments, and Greene is surely committed to claiming that such judgments must involve utterly different psychological processes than the counterintuitive utilitarian judgments on which the research has so far focused. One possibility that is in line with Greene's model is that judgments that appear to be counterintuitive deontological judgments really reflect, not the effortful overcoming of a utilitarian intuition, but unusual affective responses—e.g. an atypically strong aversion to lying.

In any case, apparent everyday examples of counterintuitive deontological judgments, and utilitarian judgments based on intuition, have so far been strangely ignored by researchers. What would we find if we examined them?

Taken at face value, Greene's dual process model makes clear predictions. Utilitarian judgments should involve deliberative processing, whether they are (in the loose sense set out above) intuitive or not. And the neural and behavioral correlates of counterintuitive deontological judgments should be utterly different from those of counterintuitive utilitarian ones such as in the Footbridge dilemma. The alternative hypothesis I've just offered makes exactly contrary predictions: Counterintuitive judgments should engage similar brain areas whether or not they are deontological or utilitarian; and a parallel prediction follows for intuitive judgments. In a recent neuroimaging study, we set out to test these competing hypotheses.

2.4. Intuitiveness vs. Content

In that study, we first operationalized "intuitiveness" by surveying the unreflective judgments of an independent sample of non-philosophers, and classifying types of judgments as intuitive if they were clearly dominant—if they were made unreflectively by 13 or more out of the 18 independent judges; the contrary judgments were classified as "counterintuitive" (Kahane et al. 2012).[14] In addition to the commonly used

[13] An even more embarrassing counterexample is provided by a recent study in which Greene himself was involved. That study found, across a wide range of decision contexts, that most subjects find cooperative decisions more intuitive than selfish ones, which typically require greater deliberative effort (Rand, Greene, and Nowak 2012). It's hard to see how this squares with Greene's model. Moreover, to the extent that an association with deliberation is supposed to give support to a normative view—as Greene often seems to assume—then these results should strongly support rational egoism.

[14] On this operationalization, "intuitiveness" is a *population-relative* notion, so judging that we are forbidden to push in Footbridge counts as intuitive in this sense even though, as we shall see in section 4, some

dilemmas such as Footbridge where the utilitarian choice is typically counterintuitive, we also used dilemmas, not previously studied, where the deontological choice is counterintuitive, such as refusing to lie to prevent harm. Importantly, although only a minority of these lay judges endorsed refusal to lie in such cases, it is highly unlikely, to put it mildly, that this refusal was based in adherence to some explicit deontological theory.

We then used fMRI to compare the neural correlates of responses to these two kinds of dilemmas. Setting aside many points of detail, our main finding can be summarized as follows: The apparent neural and behavioral differences between utilitarian and deontological judgments in trolley-like dilemmas appear to be driven almost exclusively by differences in their intuitiveness, not in their content.[15] To focus on the case that is most relevant to our discussion, "utilitarian" judgments such as that it is appropriate to push in Footbridge, and what we can call "ultra-deontological" judgments, such as refusing to lie to prevent harm, were associated with strikingly similar patterns of neural activation, compared to contrary intuitive judgments. In other words, counterintuitive judgments with *radically opposing* contents were based in *highly similar* neural processes.

These results are plainly incompatible with Greene's dual process model.[16] They show, first, that it's not the case that deontological judgments are associated with automatic processing, and utilitarian with deliberative processing, across different moral domains.[17] But our results also strongly suggest that the differences between utilitarian and deontological judgments reported by previous studies largely reflect differences in

individuals may actually lack this intuition. Notice that we don't define intuitiveness as merely a matter of consensus. The independent judges were explicitly asked for their *unreflective* response; after reflection, some gave different answers to some of these dilemmas.

[15] See Kahane et al. 2012 for the full details of the analysis that supports this conclusion. For a similar conclusion based on a different approach, see Baron et al. 2012.

[16] In an empirical response to our study, Paxton et al. 2013 report that responses to the Cognitive Reflection Test, a measure developed to measure counterintuitive thinking in a non-moral context, was associated with increased utilitarian responses both in a personal dilemma and in one of the lying dilemmas we used in Kahane et al. 2012. This is an interesting result that is not predicted by my remarks above. However, even setting aside the fact that only a single dilemma of each type was used in this study, and other methodological worries (see e.g. n. 9), this result does not really address the main result of our study: that similar neural processes were associated with counterintuitive moral judgments of opposing contents, and that the differences between utilitarian and deontological judgments within (appropriate) personal dilemmas were almost entirely explained by differences in intuitiveness.

[17] Notice that our main aim was to compare counterintuitive judgments of opposing contents. It *wasn't* to demonstrate that in some contexts, deontological judgments also involve deliberative processing. In fact we didn't find that counterintuitive judgments were generally associated with greater response times, or with activation in classical cognitive areas, compared to contrary intuitive ones. In fact we didn't find this *even* when we compared only counterintuitive utilitarian judgments with intuitive deontological ones—the comparison closest to that in Greene et al. 2004. Counterintuitive judgments were, however, associated with greater perceived difficulty, and with activation in the subgenual part of the rostral ACC, an area that has been implicated in affective conflict (Etkin et al. 2006), but also in feeling guilt (Zahn et al. 2009a; Zahn et al. 2009b). So we actually found only limited support for the hypothesis that counterintuitive judgments—*including* counterintuitive *utilitarian* judgments in trolley-like cases—are generally associated with deliberative processing.

intuitiveness even when we consider *only* "personal" dilemmas like Footbridge cases. So it seems that the neural and behavioral correlates of paradigmatic "utilitarian" judgments merely reflect deliberation to a counter-intuitive conclusion—and not anything distinctively utilitarian.[18]

2.5. A Less Exciting Dual Process Model

So what we get, instead of Greene's exciting and controversial dual-process model, is an alternative and rather less surprising dual-process model:

The Banal Dual Process Model. Intuitive judgments are generally associated with automatic processing, and counterintuitive judgments with deliberative processing.

Greene has suggested to me that his dual process theory is really meant to essentially make this more general claim about intuitive and counterintuitive judgment. This, however, is certainly not how the theory has so far been presented or understood. If Greene's claim is simply that many utilitarian judgments are counterintuitive, and that the processes involved in making such judgment merely reflect the processes generally involved in making counterintuitive judgments, then these are near truisms. It is hardly a great surprise that counterintuitive judgments are harder to make, or that they require more deliberative effort. And since we know that some utilitarian judgments are counterintuitive, it is also hardly surprising that these judgments are harder to make . . . But this would give no support to the ambitious theoretical and normative arguments that Greene and others have elaborated on the basis of the dual process model—after all, these truisms imply nothing interesting that is *specific* to utilitarianism, let alone anything *favorable* to utilitarianism.

It might be replied that even if deliberative processing is generally associated with counterintuitive moral judgments, such processing nevertheless *favors* utilitarian judgments. If this is a general claim, not specific to trolley-like cases, it's hard to see what evidence is supposed to support it. How exactly are we supposed to measure such a general tendency?[19] It is also important to distinguish this interesting sounding hypothesis from the far less exciting claim in the reverse direction: that utilitarian judgment more often involves deliberative processing *simply because* utilitarianism issues more counterintuitive conclusions compared to many other moral views. The latter claim is a mere truism. We don't need fancy neuroimaging to know *that*.

Let me summarize this section. I have argued that even if utilitarian judgments in trolley-like dilemmas are deliberative, this deliberation is merely generic. If these judgments involve deliberative processing, this is merely because they are *counterintuitive*,

[18] In line with this, we also didn't find common correlates for so-called "utilitarian" judgments across different domains.

[19] And see again the Rand, Greene, and Nowak (2012) study mentioned in n. 13, associating deliberation and selfish decisions. Does deliberative processing favor *both* utilitarianism and rational egoism?

not because they're *utilitarian*. And the processes involved are (broadly) *the same* as those involved in counterintuitive *ultra-deontological* judgments.

3. Deliberating about What?

3.1. "Utilitarian reasoning" and the Footbridge dilemma

In this section, I want to take the argument one step further.[20] Consider again the problematic inference from evidence associating "utilitarian" judgments with deliberative processing, to the conclusion that this processing reflects the distinctive utilitarian character of these judgments. To decide whether this inference is valid, we need to know what it *would mean* for such processing to have some distinctive utilitarian character. Unfortunately, however, Greene's dual process model is surprisingly unclear on this question.

Greene sometimes writes that utilitarian judgments are generated by "utilitarian reasoning." But what does Greene mean when he speaks of "utilitarian reasoning" (or "cost–benefit analysis") leading, say, to the judgment that it is morally appropriate to push the stranger in Footbridge?

Greene does not, of course, think that individuals making such judgments explicitly endorse act utilitarianism or any similar ethical theory. Neither is it plausible that they endorse such a theory implicitly, since virtually all of them make *some* deontological judgments in *some* contexts (see Kahane and Shackel 2011). Still, if they are to be said to be making "utilitarian" judgments even in the thinnest sense, their *ground* for judging that it's appropriate to push the stranger must be that this would lead to better consequences. And if they reach this conclusion by explicit reasoning, as Greene holds, then they must be reasoning from a corresponding general moral principle. This means that such individuals must be following something at least approximating the following piece of reasoning:[21]

(1) We are required to impartially maximize well-being.
(2) 5 lives > 1 life.

Therefore

(3) We are required to sacrifice 1 to save 5.

If this is what individuals are going through when they arrive at utilitarian conclusions, then they would indeed be making an inference that *sets out* from a utilitarian premise to a utilitarian conclusion—though I'll later question whether it's usefully called utilitarian *reasoning*.

[20] The argument in this section is heavily based on Kahane 2012. Some empirical and conceptual detail has been removed to clarify the argument, and several new points have been added. Still, readers already familiar with (or even persuaded by) the argument of that paper may prefer to skip to section 4.

[21] This picture is essentially endorsed in Cushman, Young, and Greene (2010).

3.2. Deliberative Processing of What?

Let's suppose for the moment that subjects are really engaged in the piece of reasoning Greene ascribes to them.[22] Greene's main claim is that when subjects engage in such reasoning and reach utilitarian conclusions, they uniquely do so using deliberative processing. This is what the empirical evidence cited above is supposed to show.

But we should ask: What exactly does this deliberative processing reflect? Let's briefly consider some possibilities.

RECOGNIZING A FOUNDATIONAL PRINCIPLE?

If the deliberative processing reflected recognition of the moral principle that we are required to maximize utility then it would perhaps reflect a distinctly utilitarian thinking. But this is highly implausible. Even if someone reached this normative view through effortful reflection when first confronted with, say, the Footbridge case, surely this is something they only need to do *once*, so this anyway couldn't account for the greater levels of deliberative processing found in responses to a large set of dilemmas.

Notice, moreover, that this is a foundational moral principle—such a principle can serve as a premise for reasoning, but it's not itself plausibly supported by any kind of inference.[23] So to the extent that the deliberative processing associated with utilitarian judgment is supposed to reflect *reasoning*, it can't reflect the non-inferential judgment (aka *intuition*) that we should maximize utility.

This example nicely illustrates the limits of thinking about moral psychology using the dual process framework. If the utilitarian principle that we ought to maximize the greater good is non-inferential, then it's in one good sense intuitive. Yet it is also claimed to be the product of "reason," rather than anything external to it (cf. Singer 2005; de Lazari-Radek and Singer 2012). Nor need it be immediate: That insight may strike someone after long, effortful reflection.

COUNTING THE NUMBERS?

Another possibility, which is often implied by the way Greene and others describe utilitarian judgment, is that the deliberative processing reflects the calculation of "5 lives > 1 life."

The first problem with this suggestion is that this *isn't* any kind of utilitarian reasoning. This is just standard *non-moral* reasoning, used to apply a moral principle—*any* kind of moral principle.[24] It's easy to vary the degree of non-moral reasoning needed to apply different moral considerations in a given context. Some "utilitarian" questions involve little or no calculation ("Should we give someone in agony a painkiller?"),

[22] See Kahane 2012, §2.3 for further reasons to doubt this.

[23] This is a point that is often emphasized by utilitarians like Singer—see Singer 2005; de Lazari-Radek and Singer 2012.

[24] This point is actually supported by Greene's own work: Shenhav and Greene (2010) report that so-called "utilitarian" reasoning has the *same* neural correlates as non-moral reasoning.

and some deontological questions require quite a bit of calculation ("Should someone break a promise to a friend that conflicts with 3 promises to his daughter, when this requires lying twice?"). Responses to such questions are likely to differ in reaction times or DLPFC activation, reflecting obvious differences in complexity. But of course these processing differences would tell us absolutely nothing about the psychology of utilitarianism or deontology. Non-moral reasoning is simply irrelevant to that issue. It follows that if the deliberative processing associated with utilitarian judgment merely reflects the calculation of "5 lives > 1 life," this would merely be a contingent artefact of the particular stimuli Greene and others are comparing, rather than any kind of interesting feature of utilitarian vs. deontological thinking.

In any event, however, it is extremely unlikely that the deliberative processing reflects this simple calculation. Non-utilitarians are not numerically challenged. It is near certain that *all* subjects considering Footbridge and similar dilemmas make this simple calculation, whether or not they reach a "utilitarian" conclusion. It is astonishing that so many researchers seem to think otherwise. There is a further problem: It is surely the case that *no* cognitive effort is required to judge that 5 is greater than 1! As we shall see in section 4, there is in fact evidence that engaging in such "utilitarian reasoning" can actually be quick and utterly effortless.

3.3. So What Does the Deliberative Processing Reflect?

The argument I have been developing so far largely converges with the empirical evidence I presented in section 2. It suggests that if utilitarian judgments are associated with deliberative processing, that processing *doesn't* reflect any kind of "utilitarian reasoning." In the previous section, I already argued that this processing really reflects a generic type of deliberation to a counterintuitive conclusion. In the rest of this section, I want to look more closely into what such deliberation might involve in the context of trolley-like dilemmas.

OVERCOMING CONFLICT?

If "utilitarian" judgments in Footbridge involve deliberative processing then that processing couldn't reflect "utilitarian reasoning." So it must reflect something else. What could that be?

Greene's own account suggests a further option. After all, his account says that individuals not only engage in such "utilitarian reasoning," but also need to actively suppress the pre-potent emotional response pushing them in the contrary deontological direction:

Deliberation reflects conflict. When deontological intuitions are present, effortful deliberative processing is needed if one is to arrive at a contrary utilitarian judgment.

This is supposed to be why we find greater activation in the dorsal Anterior Cingulate Cortex (dACC), an area associated with conflict, when subjects make utilitarian judgments. As Greene often points out, the dACC is also active in the classical Stroop

paradigm, where deliberative processing is needed to overcome the pre-potent impulse to name the color word rather than the color in which it is written (Greene et al. 2004, 390).

I have so far deliberately ignored this part of Greene's theory. Greene typically presents the deliberative processing associated with utilitarian judgment as reflecting both "utilitarian reasoning" *and* the conflict generated by the contrary emotion. The argument so far shows that it can *at most* reflect the latter.

The problem is that talk about emotional conflict and overcoming an emotional response is ambiguous. It can mean two rather different things. The first interpretation, which is implicit in Greene's view, doesn't really support his model, and is also implausible. The second, more plausible interpretation is simply incompatible with Greene's overall theory. I will consider each interpretation in turn.

RESISTING/INHIBITING A PRE-POTENT EMOTION?

If we take the analogy to the Stroop paradigm literally, then the deliberative processing should reflect the effort that individuals are making to resist a spurious distorting influence (see Paxton and Greene 2010). Such outright rejection of one's intuitions is of course common in genuine utilitarian thinking for the simple reason that utilitarians reject *all* non-utilitarian moral considerations, and thus all of the intuitions that (at least sometimes) drive them.

The first problem with this suggestion is that there is nothing distinctly utilitarian in rejecting some intuition as spurious—this is something that both utilitarians and deontologists do. Indeed, we also often dismiss our intuitions in this way in many non-moral contexts. In any case, since this is a purely negative operation, it tells us nothing *positive* about the source and nature of utilitarian thinking. Instead of the claim tying utilitarian judgment and deliberative processing, we again get the banal claim that whenever strong intuitions are present, effortful deliberative processing is needed if one is to arrive at a judgment contrary to these intuitions. It should be obvious that this near truism does not support the dual process model as such, let alone the grandiose theoretical and normative conclusions that Greene and others want to derive from it (see again Greene 2008; Singer 2005).

In any event, it turns out that this interpretation of the data is also highly implausible. If subjects rejected their aversion to intentionally killing an innocent person as expressing any kind of genuine moral reason, then they should surely reject it as such a reason *quite generally*. But few subjects make consistently utilitarian decisions across the board even in the restricted context of trolley-style dilemmas—though in section 4 we will consider subjects who at least approximate such a pattern—so when subjects endorse the deontological answer to such dilemmas, are they simply suddenly "overcome" by an emotion they take to be spurious, as people sometimes get confused and make errors in the Stroop paradigm? This is utterly implausible. Moreover, subjects tend to report feeling guilty after making utilitarian judgments in personal dilemmas (Choe and Min 2011),[25] and, when

[25] More precisely, subjects reported feeling guilty when making utilitarian judgments in an overwhelming majority of personal dilemmas. However, Choe and Min found no correlation between a trait disposition to guilt, and rates of utilitarian judgment.

given the chance, they report finding that both options are somewhat *wrong* (Kurzban, DeScioli, and Fein 2012). Neither of these results makes much sense if subjects are just resisting an insistent emotional urge which they perceive as utterly spurious.

NON-UTILITARIAN WEIGHING OF OPPOSING DUTIES

The subpersonal talk about conflict generated by a contrary emotion is also compatible with another, and more plausible, picture of the deliberation that most non-philosophers go through when they make "utilitarian" judgments in Footbridge.

In most moral dilemmas, we face genuine opposing moral reasons, and try to figure out which of these "wins" in the given context. We don't treat the reasons we ultimately reject as a mere psychological annoyance we need to repress. This is what makes moral dilemmas difficult: There are strong reasons supporting each of the conflicting choices.[26]

Once we begin to think of moral dilemmas in this more plausible way, an alternative account of what typically happens in trolley-like dilemmas naturally suggests itself. Subjects start, as I suggested earlier, with a moral reason or principle:

(1) We have reason to reduce harm (or Prima Facie *Duty to Save*).

And they count the numbers:

(2) 5 lives > 1 life.

But they also recognize that

(3) There is reason not to intentionally harm an innocent person (or Prima Facie *Duty Not to Harm*).

And at least some of them conclude, after some reflection, that

(4) In *this context*, there is *more* reason to minimize harm.

Which is why they conclude that

(5) It is permissible to sacrifice one to save 5.

If *this* is what is going on, then not only is it misleading to describe this as "utilitarian reasoning," but what we have here is really a paradigmatic example of a distinctly *deontological* form of deliberation (for a classical defence of this understanding of everyday moral deliberation, see Ross 1930). After all Act Utilitarianism leaves *no* space for such

[26] Of course, even on the utilitarian view, there is *some* reason *not* to push the fat man—needless to say, the fat man's death is a very bad consequence. But this counter-reason cannot generate conflict, or the need for effortful deliberation—on this utilitarian view, the number of lives saved in each of the two options are straightforwardly commensurable. As I have been repeating, there is neither effort nor conflict in concluding that 5 is greater than 1.

weighing of competing moral duties or reasons: On Act Utilitarianism (understood as a decision-procedure) there is only a *single* duty, viz. to maximize impartial well-being, and its application to different contexts.

3.4. Misunderstanding Moral Deliberation

Greene often portrays deontological responses as immediate judgments that some act is absolutely wrong, which he contrasts with the sophisticated utilitarian weighing of competing concerns (see e.g. Greene 2008, 64). But this gets things exactly upside down. Most plausible forms of non-utilitarian thinking (including, of course, "commonsense" morality) are not absolutist in this way. Instead, non-utilitarian moral deliberation typically involves precisely the weighing of competing moral considerations (including considerations about consequences), whereas utilitarian deliberation leaves space *only* for the (non-moral) comparison of the causal consequences of different lines of action. And, to repeat, in cases like Footbridge this utilitarian calculation should be obvious and effortless.

This more plausible alternative has so far been overlooked because current research relies on an incredibly narrow conception of what deliberation is or can be. It unthinkingly identifies deliberation with inference, and sharply contrasts it with intuition, as if inference is an utterly independent, competing source of judgment—neglecting to see that the premises for reasoning also need to come from somewhere, that effortful deliberation needn't involve anything like inference, and, moreover, that there are certain kinds of intuitions and emotions that appear *only* in deliberative contexts.

It is because of that narrow conception of deliberation that Greene and others simply assume that if deliberative processing plays a part in generating "utilitarian" judgments, then this means that this deliberative processing is generating these judgments *from scratch*, and that it does so by *inference*—as if the only alternative is for the deliberative processing to produce mere rationalization of some pre-existing emotion or intuition. But deliberative processing can play a genuine causal role in producing moral judgments without necessarily being the source of their content.

As described above, deliberative processing can help us decide between competing *pro tanto* moral reasons (or "prima facie duties"[27]) by generating an *all-things-considered* judgment about what ought to be done.[28] And that decision can be conscious and effortful without being the result of *inference*—the judgment that, in some given context, one duty *outweighs* another is almost certainly *also* based in intuition! (Notice that this would be a *second order intuition* resolving a conflict between two first order intuitions. For a classical statement of the view that we need something like intuition to resolve conflicts between opposing principles, see again Ross 1930).[29]

[27] I follow W. D. Ross in speaking of "prima facie duties," but the relevant reasons are not literally "prima facie" (merely appearing to be genuine reasons), and are best described as "*pro tanto*" (genuine reasons that can be outweighed in a given context).

[28] Nichols and Mallon (2005) show that non-philosophers can easily distinguish between whether an act violated some rule (weak impermissibility) and whether that act was wrong, *all things considered*.

[29] I'm not claiming that *no one* treats the intuition not to push as spurious, only that most people don't. There certainly are clear cases where people reject their intuitions/affective reactions as simply morally

It might be objected that subjects couldn't be going through this deliberative process because, although something like the principle of utility is an explicit principle that people actually cite, the resistance to, e.g., pushing the stranger in Footbridge is an intuition that most people are unable to articulate as an explicit principle (Cushman et al. 2006). But this is irrelevant. We can be conscious of, and torn between, opposing moral considerations even if we can't articulate them as fully explicit principles.[30]

This gives us an alternative explanation of the role of deliberative processing in utilitarian judgment:

Deliberation as weighing of opposing reasons. "Utilitarian" judgments are generated by deliberative processing reflecting deliberation about the proper weighing of various moral reasons/principles—some or all of which are deontological in character—when these generate conflicting verdicts about a given situation.

Notice that this account is perfectly compatible with the claim that "utilitarian" judgments in "personal" trolley-like dilemmas are more closely associated with deliberative processing. In many moral contexts, one of the opposing moral reasons or principles can be especially salient, and dominate, even pre-empt, deliberation. It might thus take some effort to see and give proper weight to competing considerations that are less salient.

3.5. "Utilitarian" Judgment ≠ Utilitarian Reasoning

Let me end this section by tying the above argument to a familiar objection to utilitarianism. Consider Bernard Williams's famous anti-utilitarian example of Jim and the Indians, where Jim is told that if he shoots one Indian, the lives of several others would be spared (Williams in Smart and Williams 1973). Although this is sometime misunderstood, the point that Williams wanted to make with this example *wasn't* that we *shouldn't* make the "utilitarian" decision to save more lives—he was actually inclined to think we should. Williams's point was rather that this decision shouldn't be as *easy* as is implied by Act Utilitarianism; it should be a *difficult, agonizing* choice. And indeed, unlike a minority of individuals which (as we shall see in section 4) do find this an easy choice, most people seem to find such "utilitarian" choices difficult in *exactly* the way Williams described—and not merely because they need to overcome some irritating but spurious "pre-potent response."[31]

spurious—one example might be the way that, according to Haidt, liberals regard their disgust reactions to harmless violations (Haidt 2001).

[30] In fact the utilitarian principle is itself indeterminate in multiple ways that utilitarians still disagree about.

[31] If you could put Bernard Williams in the scanner, and ask him for his view about the Jim and the Indians case, he would engage in agonizing deliberation about what one should do in such a situation. Suppose he ends up endorsing the "utilitarian" conclusion that Jim should shoot one of the Indians to save the greater number. It would be absurd to then say that, since he arrived at this conclusion through effortful deliberation, this shows that utilitarianism is based on more cognitive processing. It would be even *more* absurd to think that by studying the neural correlates of Williams's agonizing deliberation, we are learning anything

Greene et al. 2004 report that utilitarian judgments were associated with greater activation in the insula, which they speculate might reflect repugnance at the sacrifice of the one. Moretto et al. 2009 report that a strong emotional reaction (reflected in skin conductance response) *followed* utilitarian judgments. And as we saw, Choe and Min 2011 report that utilitarian judgments in most personal dilemmas were reported by subjects to be accompanied by some guilt.[32] It's hard to square these three findings with the common assumption that subjects view the deontological intuition as merely a gut reaction that needs to be resisted—but they nicely fit Williams's 1965 remarks about the *moral residue* we feel when we are tragically forced to choose one of two profoundly bad options.

The lesson here is that it is a mistake to assume that if a moral judgment is "utilitarian" (in the narrow sense that, in a given moral context, it favors the act that maximizes utility), then this implies that the deliberation that generated it must also be "utilitarian." Quite the opposite: Such "utilitarian" judgments are typically based in a distinctly deontological form of deliberation that is profoundly *incompatible* with utilitarianism.[33]

In this section, I've argued that when utilitarian judgments involve deliberation, it's not just that this deliberation doesn't reflect anything *distinctly* utilitarian—it's a form of deliberation that is resolutely *non-utilitarian* in character![34] *If* an association with deliberation is supposed to *support* a moral theory, as Greene and others think (a rather questionable assumption), then, ironically, this empirical evidence should actually support *non-utilitarian* views.

4. Cold or Calculating? "Pure" Utilitarian Judgments are Non-Deliberative

4.1. Utilitarian Judgment without Deliberation

Utilitarians hold that their counterintuitive conclusions are simply what results when, instead of just following our immediate gut reactions, we use moral reasoning to critically scrutinize them (Singer 2005; Unger 1996). This picture of utilitarian psychology

about the psychological sources of utilitarian thinking. I have been arguing in this section that a lot of current research on "utilitarian" judgment falls exactly into this absurd mistake.

[32] Utilitarian judgments in some dilemmas were more strongly associated with disgust and other emotions—but these also happen to be the dilemmas that Nick Shackel and I have previously shown to be utterly irrelevant, as now also conceded by Greene (see Kahane and Shackel 2008; Kahane and Shackel 2010).

[33] This is just one reason why it's misguided to suggest, as Greene does, that we take "utilitarian" to refer to whatever processes turn out to underlie judgments with "characteristically utilitarian content" (Greene 2008).

[34] Needless to say, if utilitarianism is understood only as a criterion of rightness (and not also as a decision procedure), then it is logically compatible with any form of deliberation—including rational egoism or applying the Categorical Imperative. But Greene's claims about the origins of utilitarian thinking *require* treating it, at least in this context, as referring to a distinctive decision procedure.

of course sits nicely with Greene's dual process model, which portrays lay people who make utilitarian judgments as thinking harder than the rest.

In the previous sections of this chapter, I have cast some doubt on this claim. In this section, I want to take the argument yet one step further. I will point to evidence that strongly suggests that there are in fact some individuals who arrive at utilitarian conclusions through a process that more closely resembles a utilitarian decision procedure. The problem is that it appears that *these* utilitarian judgments are *non-deliberative*—they are driven, *not* by greater cognitive effort, but by an *affective deficit*. Such individuals are not more "calculating" in the sense of being more rational. They are simply colder.

This alternative picture of "utilitarian" psychology is actually supported by a growing body of evidence. Some of this evidence comes from clinical populations. Patients with lesions in the vmPFC (Ciaramelli et al. 2007; Koenigs et al. 2007; Moretto et al. 2009) and with frontotemporal dementia (Mendez et al. 2005), conditions associated with deficits in empathic concern and social emotion and with disordered social behavior, exhibit increased rates of utilitarian judgment in emotionally-loaded moral dilemmas, apparently because such patients lack the normal aversive response to harming—they lack, if you want, the intuition that it's wrong to push the stranger in Footbridge.

Now this result is taken by Greene and others to be a key bit of evidence *supporting* the dual process model: If the absence of social emotion leads to an increase in utilitarian judgment, this supports the claim that deontological judgment is based in emotion (see e.g. Greene 2008). This, however, is only one half of the dual process model. What hasn't been noticed so far is that vmPFC patients at the same time raise a worry about the other, more interesting part of the model.

I have earlier argued that to the extent that "utilitarian" judgment is associated with controlled processing, that processing doesn't reflect any kind of "utilitarian reasoning." (For isn't it just obvious that it's *better* to save 5 than to save 1?) This argument actually generates a straightforward empirical prediction: that such reasoning (i.e. the psychological processes directly pushing us in the direction of pushing the fat man in Footbridge) can actually be *fast* and *effortless*. And indeed, so it is in vmPFC patients. A recent study has shown that in vmPFC patients, utilitarian judgments in personal dilemmas were associated both with weaker skin conductance responses, and with shorter reaction times, compared to healthy subjects (Moretto et al. 2009). That is, once you remove the opposing force of the deontological consideration against directly harming someone, the "utilitarian" solution becomes pretty intuitive.[35]

[35] It is also worth noting that the area of the DLPFC that Greene et al. 2004 reported to be associated with utilitarian judgment is *damaged* in many of these VMPFC patients (Moll and de Oliviera 2007). Since other parts of the DLPFC are intact, this doesn't itself show that they are incapable of controlled processing, or that they are not engaged in such processing when making utilitarian judgments. What this *does* strongly suggest is that, if the DLPFC activation reported in Greene et al. 2004 reflects the controlled processing healthy subjects engage in when they make utilitarian judgments in Footbridge, then this controlled processing *couldn't*

This finding may not seem so significant. After all, this evidence relates to rare and unusual brain damage. It tells us nothing about how ordinary, healthy individuals arrive at utilitarian judgments.

But more recent evidence shows that this pattern of response is actually far from unusual. To begin with, several recent studies report greater rates of utilitarian judgment in individuals high on antisocial traits such as psychopathy (Glenn et al. 2010; Bartels and Pizarro 2011; Koenigs et al. 2012; Wiech et al. 2013; though see also Cima et al. 2010; Glenn et al. 2009).

These findings have been overinterpreted by the press. *The Economist*, for example, reported one such study by announcing that "Goodness has nothing to do with it: Utilitarians are not nice people."[36] This is of course a blatantly fallacious inference. From the fact that some not so nice people are utilitarian, it hardly follows that all utilitarians are not nice people. This claim also relies on the terrible inference I mentioned earlier, from "utilitarian" responses to trolley-like dilemmas to general claims about utilitarian judgment. But some defenders of utilitarianism are happy to make that terrible inference when it associates utilitarian judgment with something more flattering!

However, it would *also* be a mistake to assume that these results merely relate to a tiny, abnormal minority. "Proper" psychopaths are indeed a tiny minority, but some of the studies I just cited actually examined ordinary individuals who happen to have psychopathic and similar antisocial traits—individuals who are a bit colder and harsher than the rest of us. And anyway these findings are just part of a much broader trend in recent research suggesting that utilitarian judgments in healthy individuals are often rooted in an atypically weak or even absent aversion to harming others. For example, higher rates of utilitarian judgment in normal, healthy individuals have been associated with lower rates of trait empathy (Crockett et al. 2010; Choe and Min 2011), as well as with higher levels of testosterone (Carney and Mason 2010), which often leads to reduced empathic concern (Hermans et al. 2006). Another study found that a disposition to utilitarian judgment was associated, not with greater cognitive ability, but with lower emotional intelligence, and specifically with a difficulty with reasoning thoughtfully about one's emotions (Koven 2011).

More directly, a strong "utilitarian" disposition was associated with reduced aversive reactivity to harming others, as indexed by peripheral vasoconstriction (Cushman et al. 2012), with reduced skin conductance response (Moretto et al. 2009), and, most importantly for us, with significantly lower response times (Greene et al. 2008). In other words, even in ordinary, healthy individuals, a strong disposition to utilitarian judgment in "personal" moral dilemmas is often due, not to increased critical reflection that has overcome deontological intuitions, but to the *absence* of these intuitions,

reflect a utilitarian cost–benefit analysis, since these patients can reach this utilitarian conclusion without engaging this brain area.

[36] *The Economist*, 24 September 2011.

and more generally, to reduced empathic concern. And when ordinary people are "colder" in this way, their utilitarian judgments also become pretty fast and effortless.

Stated more generally, the conclusion we should draw is the following:

Deliberative effort as function of deontological commitment. The more weight one gives to deontological constraints on harming, the more deliberative effort is needed to conclude that one should set them aside in a given moral context; the *less* seriously one takes such constraints, the more *effortless* and non-deliberative "utilitarian" judgments become.

In other words, the degree to which deliberative effort is involved in so-called "utilitarian" judgments really reflects the strength of one's *deontological* commitments! Without such commitments, "utilitarian" judgments require little or no deliberation.

4.2. Utilitarian Judgment: Hot or Cold

Let us return now to the neuroimaging study we discussed in section 1. One of our findings was that counterintuitive moral judgments in general, as well as counterintuitive utilitarian judgments in particular, were perceived as more difficult, and were associated with increased activation in the anterior cingulate cortex, and specifically in its subgenual part, compared to contrary "intuitive" moral judgments (Kahane et al. 2012). This is an area of the brain that has been repeatedly implicated in empathic concern and guilt (Decety, Michalska, and Kinzler 2012; Green et al. 2010; Moll et al. 2006; Zahn et al. 2009a; Zahn et al. 2009b). This result is in line with the finding mentioned earlier that counterintuitive utilitarian judgments in healthy individuals are typically *followed* by stronger skin conductance response (Moretto et al. 2009), indicating an emotional reaction. That reaction seems to be one of guilt, or something close, given that individuals also typically report feeling guilty when making utilitarian judgments in most personal dilemmas (Choe and Min 2011).

This tie between counterintuitive utilitarian judgments and emotion is in itself compatible with Greene's model, since these emotional responses seem to *follow* such judgments rather than cause them. But as I have argued in section 3, what these results do strongly suggest is that most individuals take their deontological intuitions at face value, as representing genuine constraints on harming others that, in some unusual cases of "life and death" decisions, might nevertheless be outweighed by strong competing considerations. This is why, for most people, such decisions appear to be accompanied by guilt, or a similar aversive response to the violation of a moral norm.

But in a separate study based on the same neuroimaging data, we found that this may not be true of all individuals (Wiech et al. 2013). To begin with, we found that higher rates of utilitarian judgment were associated with "psychoticism" (or "tough headedness"), a trait associated with lack of emotionality and empathic concern, hostility, aggression, and non-conformity to social norms (Eysenck 1976).[37] Individuals

[37] We did find that need for cognition was also associated with greater rates of utilitarian judgment, but, importantly, need for cognition and psychoticism were statistically independent (Wiech et al. 2013), suggesting they may be independent factors in utilitarian judgment. However, it is worth noting that we did not

higher on psychoticism have been shown to perceive media violence as more comical and enjoyable, and show rapid habituation to violent material (Bruggemann and Barry 2002), and reduced aversion to killing enemies in a video game (Ravaja et al. 2008). Unsurprisingly, psychoticism is also strongly correlated with psychopathy (Hare 1981; Shine and Hobson 1997), a construct often seen as lying on a continuum with psychoticism (Corr 2010; Eysenck 1992) although, unlike psychopathy, psychoticism primarily aims to capture a general dimension of personality in the normal population.

Interestingly, we found that psychoticism was at once *positively* correlated with the number of utilitarian judgments and *negatively* correlated with activation in the subgenual anterior cortex. Put together with the evidence cited here, these findings seem to suggest that although most individuals find utilitarian judgments aversive, individuals higher on psychoticism do not. We already know that vmPFC patients don't feel guilt when they violate social norms (Krajbich et al. 2009), and this is of course also true of psychopaths. In other words, it seems that individuals who are strongly disposed to utilitarian solutions to "personal" dilemmas don't just lack intuitions against directly harming others, but also do not find such harm morally problematic. They do not perceive it as a regrettable (if sometimes necessary) violation of an important moral constraint.

These "colder" individuals appear to arrive at utilitarian judgments just by considering the consequences. They are not inhibited by deontological constraints against directly harming others, or need to engage in any deliberation to decide whether some significant benefit outweighs these constraints in a given context. Unsurprisingly, these individuals also tend to more consistently endorse utilitarian solutions to "personal" dilemmas compared to other individuals, who may endorse such solutions in a few cases but not in most others.[38] The "colder" individuals consider only the consequences because they do not recognize deontological constraints against directly harming others, nor, of course, find it aversive to violate such constraints. In these respects, these individuals arrive at utilitarian judgment in ways that are more closely in line with something resembling a utilitarian decision procedure—these individuals take only consequences into account, and nothing else.[39]

find any association between need for cognition and increased activation in the DLPFC or any other marker of effortful cognition.

[38] It is hardly surprising that a colder approach to moral problems, where various deontological constraints are suppressed or ignored, will tend to lead to more "utilitarian" conclusions. This is actually a familiar feature of decision-making in emergency situations such as triage. In ordinary medical contexts, doctors and nurses don't just try to maximize the outcome as a utilitarian would—they will usually try to save the lives of even those who are gravely injured, and give some attention even to relatively minor injuries. But these deontological considerations will be brushed aside in situations of emergency, where there is an imperative to focus limited resources on those who will benefit the most. We can call such practices "utilitarian," but these imperatives are really just part of commonsense morality. This point is overlooked by Greene and others because they conflate deontology and absolutism, the view that deontological constraints must *never* be broken. They thus fail to see that many judgments that they classify as "utilitarians" are really just expressions of commonsense, non-utilitarian morality. I am grateful to the editors for these points.

[39] I described such individuals as "colder." They are colder in the sense that they exhibit little empathic concern, and don't find the idea of harming others emotionally aversive. But they needn't be entirely

Now it's not generally true that people who make utilitarian judgments aren't nice. Our study actually suggests that antisocial tendencies are the driving force behind utilitarian judgment only in some individuals (Wiech et al. 2013)—"utilitarian" judgment may have more than one source. But it does appear that many of those who are more *strongly disposed* to make utilitarian judgments, and who make such judgments in ways that more closely *resemble* a utilitarian decision procedure, really aren't that nice. In fact, these individuals are disposed to utilitarian solutions *because* they aren't so nice. The important thing, though, is that these utilitarian judgments are *non-deliberative*. These individuals may lack common emotions and intuitions, but it doesn't follow from this that they are thinking harder when presented with these moral dilemmas. They just respond to them differently.

One further lesson of these results is that we need to be more careful with the way we think about counterintuitive judgments. Claims of intuitiveness and counterintuitiveness are ambiguous. In common philosophical and everyday use, they are often, explicitly or implicitly, tied to a population. When someone says that something is counterintuitive, they are not merely reporting that it goes against their own intuitions. They are assuming that it goes against the intuitions of many or most others.

Certain utilitarian claims are often said to be counterintuitive in exactly this sense. They are claimed to go against the strong intuitions of most people. That this is so in some cases is now supported by quite a bit of empirical evidence (see e.g. Cushman et al. 2006). But it doesn't follow that when someone makes a utilitarian judgment that is counterintuitive in this sense, then this means that *this* person is going against their *own* intuitions—that they are using reason to override their first, intuitive response. Here we would be sliding to a person-relative sense of "intuitive" and "counterintuitive." In fact, a claim might be both counterintuitive in the population-relative sense, yet also be intuitive *for* some individuals. And this seems to be the case for many utilitarian judgments. These judgments reflect unusual (or absent) intuitions, not the "rational" overcoming of common intuitions.

Let me summarize this section. In section 3, I argued that when "utilitarian" judgments *are* driven by deliberation, this deliberation is *non-utilitarian*. In this section, I further argued that when individuals make *"pure"* utilitarian judgments, these judgments are *non-deliberative*! These individuals endorse counterintuitive conclusions, but these conclusions are counterintuitive only relative to the intuitions of the majority—for *these* individuals, these utilitarian conclusions are *intuitive*.[40]

unemotional. In fact, several studies have tied utilitarian judgment to feeling angry (Choe and Minn 2011; Ugazio et al. 2012; see also Plaisier and Konijn 2013 on how anger can increase tolerance of antisocial behaviour). Interestingly, Jonathan Baron, a self-avowed utilitarian, independently described moral anger as the dominant utilitarian emotion (see Baron 2011).

[40] This point can be missed if we mistakenly identify intuition (and automaticity) with emotion.

5. For the Greater Good?

In section 4, I offered a very different picture of the psychology of utilitarian judgment than that defended by Greene, and assumed by utilitarians such as Peter Singer. Utilitarianism is often presented as the reasoned and systematic generalization of natural human empathic concern (Hare 1981; Singer 1979). Yet it ironically turns out that so-called "utilitarian" judgments are often driven, not by rational reflection or "generalized benevolence" (Smart 1961), but by a *deficit* in empathic concern and an indifference to harming others.

This result might seem puzzling. In section 4, I have highlighted the tie between a strong tendency to utilitarian judgment and reduced aversion to harming others. And I claimed that if we want to study the psychological processes associated with a form of moral thinking in lay people that at least *resembles* a kind of utilitarian decision procedure, we should look at this category of individuals, who appear to just consider the consequences of their actions, and to not be inhibited (normatively and psychologically) by the nature of the violent act that they are asked to contemplate doing.

But it might be objected that such individuals aren't likely to be following a genuine utilitarian outlook. Psychopaths aren't likely to be concerned with the greater good. They are likely to be concerned only with their *own* good.

This is correct. In fact, in a further set of studies, we set out precisely to investigate whether there really is any relationship between so-called "utilitarian" judgments in trolley-like dilemmas and any kind of genuine concern for the greater good (Kahane et al. under review). We found that greater rates of "utilitarian" judgment were positively correlated not only with psychopathic tendencies and reduced empathic concern, but also with explicit endorsement of *rational egoism*—while being *negatively* correlated (albeit only marginally) with whether individuals identify with the whole of humanity. In addition, we found that rates of "utilitarian" judgment were also associated with seeing uncontroversially immoral acts in a business context (e.g. stealing money from one's company) as less wrong. We also found that although individuals higher on psychopathy do appear to show a "utilitarian" tendency in their responses to "personal" moral dilemmas, their responses to such dilemmas are in fact highly sensitive to considerations of self-interest. Thus, they show an even stronger utilitarian tendency when the "utilitarian" act would also benefit them—i.e. *they* would be among the five people saved when someone is sacrificed "for the greater good"—while if they are given the choice whether to sacrifice themselves or someone else to save five others, they overwhelmingly prefer to sacrifice another person (individuals low on psychopathy show exactly the reverse pattern of response).

Utilitarianism tells us to selflessly promote the greater good—to transcend our narrow personal concern and care about the good of all human beings, even all sentient beings, as if from the point of view of the universe. But those who tend to most strongly endorse "utilitarian" solutions to "personal" dilemmas turn out to largely care about themselves… In one obvious sense, utilitarianism and rational egoism are

diametrically opposed. But on reflection, this convergence shouldn't be so surprising. Both utilitarianism and rational egoism are teleological views, concerned with maximizing the good (in one case, the greater good, in the other, one's own good). Both often come across as rather tough-headed, unsentimental, and calculating. And both treat common moral norms, intuitions, and sentiments as no more than useful instruments, to be followed only when this leads to better consequences. Is it really so surprising that when egoists are asked for their view about a moral dilemma, and given either a deontological option or a more calculating "utilitarian" one, they would tend to prefer the latter?[41]

The results I have just described further highlight the antisocial dimension in so-called "utilitarian" judgment. You might still think that, although this is an unpleasant association, we might be able to get around it. We just need to bracket this antisocial element, and this will reveal those judgments that are truly utilitarian—those responses that are really driven by greater deliberation, and by genuine concern for the good.[42]

But don't get your hopes up. In the same series of studies, we also investigated the relation between "utilitarian" judgment and a range of views and attitudes drawn directly from the work of Peter Singer—if someone tends to think that, for example, you should push the stranger in Footbridge, would they also think that it's wrong to eat meat, or that we have obligations to people in need in developing countries? Would they think that failing to donate money to save the lives of children in poor countries is wrong in the way it is wrong to refuse to step into a pool to save a drowning child? The answer is no. We found that there is no relation between a "utilitarian" tendency in personal dilemmas and any of these marks of a genuine concern for the greater good. It's not just that individuals with such a "utilitarian" tendency don't think that, say, eating meat is wrong. They don't even find eating meat, or failing to help children in poor countries, a bit *more* wrong than those who don't have such a tendency. And importantly, this lack of relation remained *even when we controlled for the antisocial dimension in so-called "utilitarian" judgment*.[43]

[41] In line with these results, a recent study found that individuals who identify as political libertarians turn out to have lower empathic concern, identify least with humanity as a whole, and don't place as much moral weight on considerations of harm and the welfare of others (Iyer et al. 2012). Needless to say, the moral views of such libertarians are clearly very far from utilitarianism—libertarianism is in certain respects a very extreme *deontological* view! Yet it still turns out that libertarians are also *more* "utilitarian" in personal moral dilemmas. . .

[42] Conway and Gawronski (2013) suggest an interesting method to try to distinguish positive "utilitarian" tendencies from mere reduced aversion to harm. Their approach, however, doesn't address any of the issues I have been pressing in this chapter. Worse, they use a number of highly inappropriate dilemmas, thereby serious compromising their results. Their analysis strangely assumes, for example, that utilitarians would be *opposed* to abortion in a young single woman who has no income and is clearly not ready to have children.

[43] It's important to emphasize that the complaint *isn't* that people who make "utilitarian" judgments in trolley-like moral dilemmas don't explicitly endorse and follow a utilitarian theory, or even that they aren't inarticulate clones of Peter Singer. We were not expecting these people to think that we should make the extremely demanding sacrifices required by utilitarianism. But in what sense are they expressing a genuine utilitarian disposition if they don't even have a *slightly* stronger tendency to more impartial altruism compared to others? You might still insist that judging that we should push the stranger in Footbridge to save five

But these results don't threaten the argument I have developed in the previous section. They only strengthen it. As I have pointed out repeatedly, it is a mistake, and a common mistake, to treat so-called "utilitarian" responses to trolley-like dilemmas as any kind of general measure of a genuine utilitarian outlook. To repeat: This type of dilemma reflects merely one particular point where utilitarianism clashes with commonsense morality (Kahane and Shackel 2010). Worse, such dilemmas focus only on a narrow and fairly negative aspect of utilitarianism, its willingness to cause severe harm to individuals when this leads to a better outcome. They entirely ignore the more positive, altruistic, and impartial core of utilitarianism.

The thrust of my argument in this chapter is that the two main driving forces of apparently "utilitarian" judgment in such dilemmas turn out to be either a form of deliberation that is explicitly *non*-utilitarian, or a form of moral response that does, *in this context*, at least *resemble* a utilitarian decision procedure (i.e. a concern for nothing besides the consequences of actions) but actually reflects a broadly selfish and antisocial tendency. Individuals who are low on empathic concern are not concerned about the greater good, but they exhibit a more unsentimental, hard-headed, and calculating approach to moral questions, an approach which manifests itself in a more utilitarian outlook in the unusual context of "personal" dilemmas, even if not in real life.[44]

Utilitarianism is a rather peculiar moral view, with a recent history, and which never spread beyond a small minority. It is not a view that is common in the general population, so it is not surprising that we don't find much evidence for it in the lay population. If we want to study the psychological correlates of genuine utilitarianism, we should study actual utilitarians—and I think there are few of these in the subject pool of most experiments.

When Greene tells us that utilitarian judgment is uniquely based in deliberation and reasoning, this can seem like good news to utilitarians; the emerging link between utilitarian judgment and antisocial tendencies is less flattering. But true utilitarians should neither cheer the relationship between "utilitarian" judgments and "rational" deliberation, nor feel discomfort about the sinister associations between such judgments and psychopathy—for, contrary to appearances, the so-called "utilitarian" responses of non-philosophers to trolley-like dilemmas may, in the end, have rather little to do with utilitarianism.

is appropriately called a "utilitarian" judgment. But judging that we should help the elderly lady across the street is also a "utilitarian" judgment in that sense. Yet studying the neural correlates of the latter judgment would surely tell us nothing at all about utilitarianism qua ethical outlook.

[44] Some philosophers with strong utilitarian inclinations have actually confessed to me that they see themselves as low empathy, unsentimental types. It is an open empirical question whether most proper utilitarians start out as high empathy types that, through a process of reflection, come to reject various intuitions, or whether they are low empathy types who are attracted, at an abstract level, to the systematic, tough-headed nature of utilitarianism, without feeling much natural concern for the welfare of others.

6. Conclusion

According to Greene's dual process model of moral judgment, our moral psychology is divided between opposing forces: an emotional, intuitive deontological half, and a rational, deliberative utilitarian half.

In this chapter, I have argued that this provocative picture of our moral psychology is mistaken. I first argued that there is in fact nothing distinctively utilitarian about deliberation to "utilitarian" conclusions: They merely reflect generic deliberation to a counterintuitive conclusion. I then went further and argued that such deliberation to "utilitarian" conclusions in trolley-like dilemmas is in fact *non*-utilitarian in character. It involves, not "utilitarian reasoning," but the weighing of opposing moral principles, a process that itself relies on intuition, at a second order level. I next argued that when individuals *do* arrive at counterintuitive conclusions in a way that might resemble utilitarian reasoning, these judgments actually turn out to be *non-deliberative*—and to reflect a strong antisocial, self-centered tendency. For these individuals, utilitarian solutions to trolley-like dilemmas don't require deliberation, because for them such solutions are actually *intuitive*. Finally, I presented evidence that strongly suggests that there is actually little connection between so-called "utilitarian" responses to trolley-like dilemmas and genuine concern for the greater good, *even* when we set aside the antisocial dimension that drives some of these responses. This shouldn't be surprising: that someone is willing to dismiss (or otherwise discount) a single specific deontological constraint against harming others in highly unusual contexts hardly shows that this person has any special concern for the greater good in any other context, let alone that they have the slightest inclination to make (or even just endorse) the demanding sacrifices required by a genuine utilitarian outlook.

Over the past decade or so, an ever growing empirical literature has been devoted to studying so-called "utilitarian" judgment in trolley-like dilemmas. The argument of this chapter is that much of this literature has been based on a series of mistaken assumptions—mistaken assumptions about the nature of the relation between deliberation and intuition, about what "utilitarian reasoning" might involve, and what moral deliberation is actually like, about the different senses in which a judgment might be intuitive or counterintuitive, about what it is to genuinely face a moral dilemma, and finally, about what it even means to make a "utilitarian" judgment.

We should welcome the use of philosophical concepts, distinctions, and examples in the cognitive science of ethics, and it is only natural that when such concepts and distinctions are adapted to an empirical setting, they may change some of their meaning, or lose some of their philosophical precision. We should even welcome bold attempts to draw ambitious normative conclusions from such empirical research. But the distinctions that philosophers draw in their armchairs often have important psychological implications, and when these distinctions are overlooked or ignored, empirical research can be led into error. Perhaps worse, when empirical research assumes problematic or even simply mistaken understandings of basic philosophical notions, it is led to overlook important possibilities, and to ignore key questions.

Dual process models may be the height of fashion at the moment, but trying to divide everything in the mind into "automatic" or "controlled" isn't any kind of advance. Nor is it useful to crudely oppose intuition and deliberation, or emotion and reason. We already have a richer, and more precise, vocabulary to describe the different ways in which we form our moral judgments. Moral deliberation often draws on intuition, and it isn't always, or even often, a kind of inference. In fact, moral deliberation not only often takes its starting point from intuition, it sometimes actually requires intuition to reach its *conclusion*—yes, there are intuitions (and for that matter emotions) that *only* make sense in the context of deliberation. Current empirical research hasn't even begun to explore this territory.[45]

Acknowledgments

This chapter draws on prior work, including Kahane and Shackel 2010; Kahane et al. 2012; Wiech et al. 2012, and especially the argument outlined in Kahane 2012—the present chapter reprises and develops that earlier argument, and then takes it several steps forward. I would like to thank my collaborators on some of these earlier papers, and especially Nick Shackel and Katja Wiech. I am also extremely grateful to the editors for excellent comments on an earlier draft of this chapter. I presented much of the material in this chapter at the workshop on *Moral Psychology and Human Agency* at the University of Michigan, as well as at events in Bielefeld and Munich. I am also grateful to the audiences in these events for many useful suggestions. Work on this chapter was supported by a University Award from the Wellcome Trust (WT087208MF) and by a grant from the Volkswagen Foundation.

References

Baron, J. 2011. Utilitarian emotions: Suggestions from introspection. *Emotion Review*, *3*(3), 286–7.

Baron, J., Gürçay, B., Moore, A. B., and Starcke, K. 2012. Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese*, *189*(1), 107–17.

Bartels, D. 2008. Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381–417.

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

[45] Once we start to think about deliberation in this richer way, numerous interesting empirical questions come into view: How are different moral (and other practical) reasons represented, aggregated, and integrated in the brain? How does our cognitive system keep tab both of different kinds of reasons (e.g. relating to beneficence, fidelity, justice, etc.) and of whether they support or oppose some particular option in a choice situation? (After all reasons of different types can simultaneously support and oppose a given option.) Does the brain represent the number of reasons in favor of some option independently of their strength? (A single weighty reason can outweigh numerous weak ones.) How does the brain represent and process the distinction between *pro tanto* reasons and all-things-considered ones (or "oughts")? Is the outcome of deliberation merely a function of the context-invariant strength of each reason, or is a function of the holistic constellation of reasons in each particular situation, as Ross and Dancy argue?

Berker, S. 2009. The normative insignificance of neuroscience. *Philosophy and Public Affairs*, *37*, 293–329.

Bruggemann, J. M., and Barry, R. J. 2002. Eysenck's P as a modulator of affective and electrodermal responses to violent and comic film. *Personality and Individual Differences*, *32*, 1029–48.

Carney, D. R., and Mason, M. F. 2010. Decision making and testosterone: When the ends justify the means. *Journal of Experimental Social Psychology*, *46*, 668–71.

Choe, S. Y., and Min, K.-H. 2011. Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making*, *6*, 580–92.

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

Cima, M., Tonnaer, F., and Hauser, M. D. 2010. Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience*, *5*, 59–67.

Conway, P., and Gawronski, B. 2013. Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–35.

Corr, P. J. 2010. The psychoticism–psychopathy continuum: A neuropsychological model of core deficits. *Personality and Individual Differences*, *48*, 695–703.

Crockett, M. J., Clark, L., Hauser, M. D., and Robbins, T. W. 2010. Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*, 17433–8.

Cushman, F., Gray, K., Gaffey, A., and Mendes, W. B. 2012. Simulating murder: The aversion to harmful action. *Emotion*, *12*(1), 2–7.

Cushman, F., Young, L., and Greene, J. D. 2010. Our multi-system moral psychology: Towards a consensus view. In J. M. Doris (ed.), *Oxford Handbook of Moral Psychology*. Oxford: Oxford University Press.

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

Decety, J., Michalska, K. J., and Kinzler, K. D. 2012. The contribution of emotion and cognition to moral sensitivity: A neurodevelopmental study. *Cerebral Cortex*, *22*(1), 209–20.

Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., Hirsch, J. 2006. Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, *51*, 871–82.

Evans, J. 2007. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–78.

Eysenck, H. J. 1976. *Psychoticism as a dimension of personality*. London: Hodder & Stoughton.

Eysenck, H. J. 1992. The definition and measurement of psychoticism. *Personality and Individual Differences*, *13*, 757–85.

Foot, P. 1978. The problem of abortion and the doctrine of double effect. In Foot, *Virtues and Vices*. Oxford: Blackwell.

Frederick, S. 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Hermans, E. J., Putman, P., and Van Honk, J. 2006. Testosterone administration reduces empathetic behavior: A facial mimicry study. *Psychoneuroendocrinology*, *37*, 859–866.

Glenn, A. L., Koleva, S., Iyer, R., Graham, J., and Ditto, P. H. 2010. Moral identity in psychopathy. *Judgment and Decision Making*, *5*, 497–505.

Glenn, A. L., Raine A., and Schug R. A. 2009. The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry*, *14*, 5–9.

Green, S., Lambon Ralph, M. A., Moll, J., Stamatakis, E. A., and Zhan, R. 2010. Selective functional integration between anterior temporal and distinct fronto-mesolimbic regions during guilt and indignation. *Neuroimage*, *54*, 1720–6.

Greene, J. D. 2008. The secret joke of Kant's soul. In W. Sinnott-Armstrong (ed.), *Moral Psychology: The Neuroscience of Morality*, pp. 35–79. Cambridge, MA: MIT Press.

Greene, J. D. 2009. Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, *45*, 581–4.

Greene, J. D., and Haidt, J. 2002. How and where does moral judgment work? *Trends in Cognitive Science*, *6*, 517–23.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., and Cohen, J. D. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*, 1144–54.

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

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–7.

Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–34.

Haidt, J. 2007. The new synthesis in moral psychology. *Science*, *316*: 998–1002.

Haidt, J. 2012. *The Righteous Mind*. New York: Pantheon.

Hare, R. M. 1981. *Moral Thinking*. Oxford: Oxford University Press.

Huebner, B., Dwyer, S., and Hauser, M. 2009. The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*(1), 1–6.

Iyer, R., Koleva, S., Graham, J., Ditto, P., Haidt, J. 2012. Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, *7*(8): e42366. doi:10.1371/journal.pone.0042366

Kahane, G. 2012. On the wrong track: Process and content in moral psychology. *Mind and Language*, *25*(5), 519–45.

Kahane G., Everett, J., Earp, B., Farias, M., and Savulescu, J. Under review.

Kahane, G., and Shackel, N. 2008. Do abnormal responses show utilitarian bias? *Nature*, *452*, 7185.

Kahane, G., and Shackel, N. 2010. Methodological issues in the neuroscience of moral judgment. *Mind and Language*, *25*, 561–82.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., and Tracey, I. 2012. The neural basis of intuitive and counterintuitive moral judgement. *Social, Cognitive and Affective Neuroscience*, *7*, 393–402.

Kant, I. 1797/1966. On a supposed right to lie from philanthropy. In M. J. Gregor (ed.), *Practical Philosophy*, pp. 611–15. Cambridge: Cambridge University Press.

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

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., and Damasio, A. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908–11.

Koven, N. S. 2011. Specificity of meta-emotion effects on moral decision-making. *Emotion*, *11*(5), 1255–61.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., and Camerer, C. F. 2009. Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *Journal of Neuroscience*, *29*, 2188–92.

Kurzban, R., DeScioli, P., and Fein, D. 2012. Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, *33*, 323–33.

de Lazari-Radek, K., and Singer, P. 2012. The objectivity of ethics and the unity of practical reason. *Ethics*, *123*, 9–31.

McGuire, J., Langdon, R., Coltheart, M., and Mackenzie, C. 2009. A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, *45*(3), 577–80.

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

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. 2006. Human fronto–mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences*, *103*(42), 15623–8.

Moll, J., and de Oliveira-Souza, R. 2007. Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences*, *11*, 319–21.

Moore, A. B., Clark, B. A., and Kane, M. J. 2008. Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*, 549–57.

Moretto, G., Làdavas, E., Mattioli, F., and di Pellegrino, G. 2009. A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*(8), 1888–99.

Nichols, S., and Mallon, R. 2005. Moral dilemmas and moral rules. *Cognition*, *100*(3), 530–42.

Paxton, J. M., Bruni, T., and Greene, J. D. 2013. Are "counter-intuitive" deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). *Social Cognitive and Affective Neuroscience*, nst 102.

Paxton, J. M., and Greene, J. D. 2010. Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, *2*, 511–27.

Paxton, J. M., Ungar, L., and Greene, J. D. 2012. Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163–77.

Plaisier, X. S., and Konijn, E. A. 2013. Rejected by peers—Attracted to antisocial media content: Rejection-based anger impairs moral judgment among adolescents. *Developmental Psychology*, *49*(6), 1165.

Rand, D. G., Greene, J. D., and Nowak, M. A. 2012. Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–30.

Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., and Keltikangas-Jarvinen, L. 2008. The psychophysiology of James Bond: Phasic emotional responses to violent video game events. *Emotion*, *8*, 114–20.

Ross, W. D. 1930/2002. *The Right and the Good*. Oxford: Oxford University Press.

Shenhav, A., and Greene, J. D. 2010. Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, *76*, 667–77.

Shine, J., and Hobson, J. 1997. Construct validity of the Hare Psychopathy Checklist, Revised, on a UK prison population. *Journal of Forensic Psychiatry and Psychology*, *8*, 546–61.

Singer, P. 1979. *Practical ethics*. Cambridge, New York: Cambridge University Press.

Singer, P. 2005. Ethics and intuitions. *Journal of Ethics*, 9, 331–52.

Smart, J. J. C. 1961. *An Outline of a System of Utilitarian Ethics*. Melbourne: Melbourne University Press.

Stanovich, K. 1999. *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum.

Suter, R. S., and R. Hertwig. 2011. Time and moral judgment. *Cognition*, *119*(3), 454–8.

Terbeck, S., Kahane, G., McTavish, S., Levi, N., Savulescu, J., Cowen, P., Hewstone, M. 2013. Beta-adrenergic blockade reduces utilitarian judgment. *Biological Psychology*, *92*(2), 323–8.

Thomson, J. J. 1976. Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204–17.

Ugazio, G., Lamm, C., and Singer, T. 2012. The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion*, *12*(3), 579.

Unger, P. 1996. *Living High and Letting Die: Our Illusion of Innocence*. Oxford: Oxford University Press.

Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., and Tracey, I. 2013. Cold or calculating? Reduced activity in the subgenual cingulate reflects decreased aversion to harming in counterintuitive utilitarian judgment. *Cognition*, *126*, 364–72.

Williams, B. A. O. 1965. Ethical consistency. *Proceedings of the Aristotelian Society* (Supplement), *39*, 103–24.

Williams, B. A. O. 1973. A critique of utilitarianism. In J. J. C. Smart and B. A. O Williams (eds.), *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Youssef, F. F., Dookeeram, K., Basdeo, V., Francis, E., Doman, M., Mamed, D., Maloo, S., Degannes, J., Dobo, L., Ditshotlo, P., and Legall, G. 2012. Stress alters personal moral decision making. *Psychoneuroendocrinology*, *37*, 491–8.

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., and Grafman, J. 2009a. The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex*, *19*, 276–83.

Zahn, R., de Oliveira-Souza, R., Bramati, I., Garrido, G., and Moll, J. 2009b. Subgenual cingulate activity reflects individual differences in empathic concern. *Neuroscience Letters*, *457*(2), 107–10.