On the Cognitive (Neuro)science of Moral Cognition: Utilitarianism, Deontology, and the "Fragmentation of Value"

Alejandro Rosas

Abstract Scientific explanations of human higher capacities, traditionally denied to other animals, attract the attention of philosophers and other workers in the humanities. They are often viewed with suspicion and skepticism. Against this background, I critically examine the dual-process theory of moral judgment proposed by Greene and collaborators and the normative consequences drawn from that theory. I believe normative consequences are warranted, in principle, but I propose an alternative dual-process model of moral cognition that leads to a different normative consequence, which I dub "the fragmentation of value" (Nagel. Mortal questions. Cambridge: Cambridge University Press; 1979). This alternative model abandons the neat overlap between the deontological/utilitarian and the intuitive/reflective divides. Instead, we have both utilitarian and deontological intuitions as equally fundamental and partially in tension. Cognitive control is sometimes engaged during a conflict between intuitions. When it is engaged, the result of control is not always utilitarian; sometimes it is deontological. I describe in some detail how this version is consistent with evidence reported by many studies and what could be done to find more evidence to support it.

Keywords Cognitive control • Dual-process theory • Evolution • Intuition • Moral cognition • Moral dilemmas • Reaction times • Value pluralism

1 Introduction

Is neuropsychological research into moral judgment [1, 2] of any relevance for the humanities and the social sciences? I merge the latter two areas of knowledge because both have, presumably, an interest in understanding human morality, religiosity, aesthetic sensitivity, shared intentionality [3], and other traits widely held to

A. Rosas (✉)
Philosophy Department, National University of Colombia,
Kra. 30, #45-03, Bogotá, Colombia
e-mail: arosasl@unal.edu.co

A. Ibáñez et al. (eds.), *Neuroscience and Social Science*,
DOI 10.1007/978-3-319-68421-5_9

be uniquely human. The understanding they seek is not primarily explanatory and scientific. Most often, they want to know what ideals, values, and human characteristics are worth preserving and promoting. And sometimes, this interest leads them to reject scientific explanations as altogether irrelevant to concerns about values.

Our initial question can be reformulated in this way: Can we draw normative conclusions from neuropsychological theories? Can they legitimately make recommendations about what morality to accept, what type of state and government to prefer, and which laws to vote for in parliament?

A vast majority of philosophers and humanists more or less intuitively, more or less reflectively, deny any normative relevance to neuroscience. As a philosopher, I belong in the heretical (albeit growing?) minority that is open to the possibility of its normative relevance—including in this openness other empirical sciences dealing with mind and morals. If by looking at sciences like psychology, cognitive neuroscience, and evolutionary biology, we come to understand what morality is, we might get a deeper grasp of its functions and peculiar authority.

In this chapter, I discuss how normative conclusions can follow from neurocognitive research into moral judgment and how they depend, crucially, on the theoretical interpretation of the data. First, in Sect. 2, I briefly reconstruct Greene's argument [4, 5] for his normative conclusion. I concisely describe the dual-process theory of cognition, its application to moral cognition, and the evolutionary presuppositions that support the normative conclusion. In Sect. 3, I present the new data on reaction times (RTs); and in Sect. 4, I describe data from cognitive load studies suggesting an alternative version of the model. Briefly, we have both utilitarian and deontological intuitions, which are sometimes in agreement and sometimes deeply in conflict. Section 5 introduces the concepts of variable utilitarian and deontological sensitivities and explains how conflict intensity varies among individuals, some of whom might also exhibit severe weakness in one or both sensitivities. The alternative dual-process theory is presented in Sect. 6. In Sect. 7, I draw the normative conclusion.

2 Greene's Normative Claim

Greene [4, 5] complemented Greene et al.'s dual-process theory of moral judgment [1, 2]—a theory that belongs within cognitive neuroscience—with a normative claim recommending utilitarianism over deontology. His collaborative neurocognitive research had shown that utilitarian responses to moral dilemmas are connected to executive decision-making, whereas deontological ones are intuitive, automatic, and emotional. He combined this finding with the idea that deontology comprises principles of action that evolved as adaptive intuitions among our evolutionary ancestors. These intuitions, however, may produce maladaptive behavior in rapidly changing, social environments [4]. Utilitarianism corrects for these maladaptive effects. It is slow and thus inefficient when quick decisions are called for, but it is flexible and adapts rationally to varying circumstances. Initially, Greene cautiously presented this normative claim as hypothetical, as an example of how neuroscience (complemented with cognitive science and evolutionary biology) can affect our normative views [4]. Since then, he developed the theory to back up this normative claim [5]. If any, this is a serious normative conclusion to draw from research in cognitive neuroscience.

Although I am not convinced of the soundness of this normative conclusion, I must emphasize I see nothing logically or scientifically wrong with the underlying reasoning. If the neurocognitive data were as Greene and collaborators presented them in their two early papers, the normative conclusion Greene inferred would be a serious contender for the truth. But the devil is in the details (of the data). The data reported by Greene et al. [1, 2] certainly seem to support a theoretical identification of deontology with intuitive, automatic thinking on the one hand and utilitarianism with controlled, reflective, effortful thinking on the other. With additional scientific premises (widely accepted among scientists dealing with mind and morals), these data enter into an argument with the following logical structure:

1. There is a difference between automatic (intuitive) and controlled (reflexive) cognitive processes (dual-process theory in cognitive science) [6, 7].
2. Automatisms evolve to deliver fast, reliable, and therefore efficient responses. But speed is traded-off against flexibility and accuracy (a constraint in the design of organisms shaped by natural selection).
3. Controlled processes correct for inaccuracies of automatic ones (hypothesis about the function of executive control) [8].
4. In evolutionary novel situations, like those that often arise when organisms live in a complex social world, controlled processes often override—and ought to override—the fast, automatic, and intuitive responses, to keep behavior in target.
5. Deontological judgments about cases are intuitive, automatic, emotional, and fast. In contrast, utilitarian judgments are controlled and slow and work to correct intuitive judgments (the brain imaging and reaction time data from Greene et al. [1, 2, 9] interpreted in the light of dual-process theory).
6. Conclusion: Deontology ought to be overridden by utilitarianism when they conflict.

Against the scientific background of dual-process theory and evolutionary biology, Greene interprets the neurocognitive results as inviting us to endorse utilitarianism. My doubts arise in regard to premise no. 5 in the above argument, namely, the neat allocation of deontological principles to evolutionary ancient and automatic processes on the one hand and of utilitarian responses (hereafter UR) to executive or cognitive control correcting intuitive and inaccurate judgments on the other. The data strongly suggest an alternative interpretation. They could point to a different dual-process theory, where not only utilitarian but also deontological responses to moral dilemmas can claim a noble origin in the executive functions.

3 Enigmatic Reaction Time Data

According to the neuroscientific evidence reported by Greene et al. [1, 2], deontological judgments activate emotional circuits in the brain, whereas utilitarian judgments activate preferentially the dorsolateral prefrontal cortex, associated with cognitive control. Additionally, behavioral data—specifically, the RT of participants confronted with personal dilemmas—show that these are longer for UR [1], a fact that also suggests the same interpretation in terms of dual-process theory. Therefore, only utilitarianism is connected to reasoning and executive functions; deontology, in contrast, is emotional, intuitive, fast, and automatic.

The idea that deontological judgments are intuitive, automatic, and emotional is quite a challenge to the traditional philosophical view linking deontology exclusively to reason, as in Kant [10]. But new evidence alerts us against overhasty claims on this point. The new evidence came primarily from corrected measurements of RT. In the course of this chapter, I also review data coming from new cognitive load studies that support a revision of Greene et al.'s original dual-process model. The neat allocation of deontological principles to evolutionary ancient and automatic processes, on the one hand, and of UR to executive or cognitive control, on the other hand, is not as promising as it seemed to be initially. As for the fMRI data, at the end of Sect. 6, we shall see that the alternative version of the dual-process theory recommends a new design for data collection.

The original evidence suggesting a difference between the RT of deontological and UR turned out to be an artifact of including inadequate dilemmas in the battery used for testing [11–13]. Greene conceded in his reply to McGuire et al.: "The apparent RT effect was generated by the inclusion of several "dilemmas" in which a personal harm has no compelling utilitarian rationale. These dilemmas reliably elicited fast, disapproving judgments, skewing the data" [14, p. 582]. However, Greene was already aware of the problem, thanks to a personal communication with Liane Young. He reacted conducting with his collaborators a new study [9] and run the analyses only on "high-conflict" personal dilemmas. This subgroup of dilemmas does have the required structure, pitting deontological against utilitarian considerations. Greene and collaborators measured the RT for utilitarian and deontological responses in two conditions: with and without cognitive load (the load was detecting the number 5 in a row of numbers scrolling across the screen beneath the dilemmas during the deliberation time). Their results show that RT increases in the load compared to the no-load condition, but solely for the UR. Load had no effect on the RT of deontological responses. This is plausibly interpreted as implying that utilitarian, but not deontological, responses use working memory resources that are being interfered with in the load condition.

Their 2008 experiment also threw one further interesting result. In a follow-up analysis, they allotted participants to two subgroups regarding their tendency (high or low) to deliver UR. The high tendency group exhibited a surprising pattern: in the no-load condition, their UR had significantly shorter RT than their deontological responses (5350 ms vs. 6070 ms, respectively; see Fig. 1, left). On the other hand,

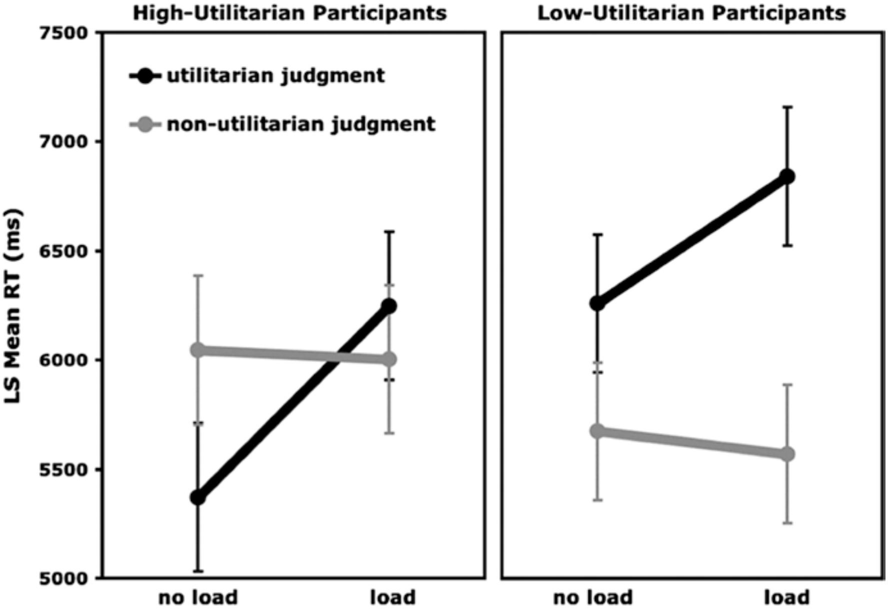

Fig. 1 Effects of load on RT for high-utilitarian (n = 41) and low-utilitarian (n = 41) groups. Original in [9, p. 1150]. Reproduced here with permission

only their utilitarian, but not their deontological, RTs were affected by load. But, precisely under load, the mean RT of their UR was not significantly higher than the RT of their deontological responses (6250 ms vs. 6000 ms, respectively; see Fig. 1, left), suggesting that some cognitive control also underlies deontological responses. So, despite the impressive result obtained comparing the load and no-load conditions, these findings about the RT are bizarre and should caution us not to endorse the dual-process theory in its original form without further investigation. Greene and collaborators grant that accounting for this result "will require a significant expansion and/or modification of our dual-process theory" [9, p. 1152].

4 Modifying the Dual-Process Theory of Moral Cognition

Greene et al. [9] ranked participants from high to low by their percentage of UR to the set of high-conflict dilemmas and divided the sample into high- and low-tendency utilitarian participants. The concept of a "tendency" to deliver UR is interesting. It could easily lead to a very different dual-process theory. A high-tendency utilitarian participant is prone to give UR easily, but deontological responses only with some difficulty. Taken to the limit, considering, e.g., only the top ranks among the high-tendency utilitarians, the "easiness" could mean that they deliver fast and intuitive utilitarian responses. Conversely, one could rank participants by percentages of deontological responses from the highest to the lowest; at the low end, we would find participants that deliver deontological responses as products of a slow, controlled, deliberative process. We would then have to admit two further types of moral judgments, impossible in the present version of the dual-process model of moral cognition but perfectly possible according to common sense: intuitive utilitarian judgments and reflective deontological judgments.

The resulting four types of judgments deliver a much messier, and less catchy, picture than the hypothesis Greene et al. proposed. This messy picture is compatible with the new evidence debunking the claim that URs have longer RT than deontological responses. Statistically, this follows from a comparison of the mean RT of both types of responses, which in this case yields no significant difference. Usually, this suggests that the RT ranges from low to high in both response types. Take, for example, a high-conflict dilemma for which the proportions of utilitarian to deontological responses are nearly equal, like *crying baby* (53.66% utilitarian response in [2]). The average RT for deontological responses ($n = 19$) is 6274 ms (range: 3199–14,445 ms). The average RT for UR ($n = 22$) is 6365 ms (range: 2453–12,456 ms) (data from [2]).[1] In principle, these data are compatible with the idea that some URs are intuitive and some reflective and the same for deontological responses. The intuitive/reflective divide would not overlap with the deontological/utilitarian divide. Reaction times alone cannot prove this, but they do suggest it. One issue raised by this possibility is this: how shall we interpret people who give intuitive deontological responses to dilemmas where the majority response is utilitarian (like impersonal dilemmas or dilemmas where killing one saves millions) or who give intuitive UR to dilemmas where the majority response is deontological (like *footbridge*)? In labeling them “intuitive,” I mean delivered without conflict. What explanation could this have in terms of the moral perspective of those participants? I shall return to this question in Sect. 5, where I shall comment on the implications of individual variation disclosed in research with moral dilemmas.

Utilitarian intuitions seem to be present in participants responding to moral dilemmas. This has been suggested in a number of studies [15–18]. Some of these studies find in moral cognition signs of intuitions as placeholders for logical operations, a phenomenon observed also in reasoning tasks [19, 20]. Additionally, one paper [21] has produced experimental evidence that deontic responders faced with impersonal dilemmas (like *trolley*) do detect a conflict with utilitarian principles, despite responding deontologically. In a follow-up paper, Bialek and De Neys report that deontic responders detect conflict in an intuitive way, because the detection is not affected by load [15]. This suggests that awareness of the conflict between utilitarian and deontological principles is itself intuitive, not an effect of a controlled process. The conflict arises from the simultaneous activation of deontological and utilitarian intuitions, implying a critique of the classic default-interventionist dual-process model. In the latter, the conflict occurs between an intuitive deontological and a controlled-utilitarian process, such that only URs qualify as controlled. In the so-called hybrid dual-process model [15], the conflict occurs between two

[1] Thanks to Josh Greene for sharing the data.

intuitions. One could presume that subjects who detect a conflict give a reflective, cognitively controlled response, independently of the type of response; but at the present state of research, this can only be conjectured rather than asserted. After all, detecting a conflict is not the same as reasoning one's way out of it.

It has also been argued that dilemmas featuring extraordinary kill-save ratios, i.e., when the ratio of lives lost to lives saved is very low—e.g., kill one to save thousands—facilitate *intuitive* UR. As evidence for this claim, Trémolière and Bonnefon report that extraordinary kill-save ratios (<1:500) influence the percentage levels of UR independently of simultaneous cognitive load of the subjects solving a dilemma task [18]. Apparently, the influence of these ratios on UR occurs intuitively, not mediated by working memory. Thanks to the pioneering research of Greene et al., we also know that impersonal[2] harm drastically increases the percentage of UR. Does impersonal harm influence the response intuitively? Moore et al. [13] showed that working memory capacity does not affect increase in UR if killing is impersonal, suggesting that this feature is intuitively processed and applied to judgment with no demand on working memory.

From a commonsense perspective, we can easily conceive of intuitive UR, contradicting the default-interventionist model. Consider the cases where the utilitarian and the deontological intuitions converge on the same action, like in *preventing the spread*. Here a doctor decides to administer a deadly poison to a person who is malevolently planning to spread HIV. This dilemma (modified to make harm nonlethal) was classified in Kahane et al. [17] as "utilitarian intuitive," and indeed most participants choose the utilitarian option when judging the appropriateness of sacrificing a victim who is about to commit a criminal action. In one of Greene's classic studies, 40 from 41 participants delivered the UR to this (unmodified) dilemma in an average RT of 4646 ms (range: 2398–12,006 ms) (data from [2]). But note that we could interpret the doctor's action as third-party punishment, which is also seen as a deontological (retributive) moral attitude. Usually, malevolent people who draw pleasure from harming others are punished in order to prevent them from harming more people, among other reasons. Doing so deters future violations, generating a benefit to the group. Arguably, we face here a paradigmatic case of the partial overlap of utilitarianism and deontology. It works like this: a regard for the good of others (one's group) bans all those actions where harm to (innocent) others is used as a means to obtain selfish benefits. Disregard of this ban leads to punishment. Justice is thus born.

Another candidate for a congruent case is telling a white lie [17]. Most subjects choose to tell a lie when the truth would cause harm unnecessarily. Note, however, that *white lie* can also be read as presenting a conflict between deontological duties—"Tell the truth" vs. "Do not harm innocent people." And yet, it is plausible to claim that people prioritize the duty not to harm in this case, because it also makes utilitarian sense. It is, perhaps, a case where utilitarian and deontological intuitions are congruent.

[2] Impersonal harm is typically unintended and committed without exerting muscular force. In Sect. 6 we discuss these two aspects separately.

In dilemmas like *white lie* and *preventing the spread*, utilitarianism and deontology support the same action. It seems that the good of others (the group) is at the root of some deontological intuitions. The good of the group requires us to constrain our freedom, ultimately in attention to the welfare of the group to which we belong. These constraints are the deontological norms.

So far so good, but this is not the whole story. Congruent cases in no way deny that many moral dilemmas present a real conflict. Utilitarianism not only prescribes justice, i.e., it not only prohibits taking away from others what is theirs: their freedom, personal integrity, belongings, and reputation. Utilitarianism also requires us to give to others what is legitimately ours when others need it urgently and to give without the framework of reciprocity that usually characterizes cooperative helping. Here deontology and utilitarianism are in tension. Sacrificial dilemmas like *footbridge* bring this tension to its utmost level, because they present cases where somebody who is not doomed or guilty is forced without consent to offer his life in sacrifice for the lives of several others. This extreme form of utilitarianism is repugnant to many people. Nonetheless, when both moralities genuinely conflict, special circumstances like harm occurring unintended or extraordinary kill-save ratios [22–24] favor UR intuitively, while yet other circumstances might influence UR through controlled processes, as we shall see in Sect. 6.

5 Individual Variation in Moral Sensitivity

In the preceding section, we encountered the construct "tendency to deliver utilitarian responses." This construct was supported with a model to predict RT by Baron et al. [25] and Baron and Gürçay [26]. They modeled the probability of a UR to a given dilemma as a function of the individual ability to give UR and of the degree of difficulty of the particular dilemma. They further argued that when ability matches difficulty, the probability is 0.5 and RTs are longest. The situation is, in their opinion, analogous to the probability of giving the correct meaning of a word depending on individual word competence and word difficulty [25]. But these cases are also different in one important respect. In moral dilemmas, identifiable objective features affect the probability of an UR (e.g., death as unintended side effect, or the kill-save ratio). These objective features have to be included in the theory and in the model. In the case of word competence, there are no such features and hence the difference.

The features that affect the difficulty or easiness of a dilemma speak always to the opposition between two sensitivities in individuals: sensitivity to utilitarian considerations and sensitivity to deontological considerations. There is a complex dynamics between these two sensitivities. First, they are not always opposed to each other. In some cases, they converge on the same response. The clearest cases of convergence are the congruent cases [21, 27]. Other less obvious cases may also favor convergence: e.g., cases of punishment and white lies, as discussed above. But when these sensitivities conflict instead of synergizing, it is possible to point to objective circumstances whose presence/absence increases/decreases the probability of an UR. When we limit our scope to sacrificial dilemmas, circumstances whose presence or absence matters are:

the death caused by the maximizing action is not intended [1, 2];
or the victim would die anyway [13, 28, 29];
or extraordinary kill-save ratios [18, 22–24];
or the victim is guilty [29];
or the agent is among the saved [13, 29];
or none of the previous, but the victim is sacrificed without exerting muscular force [5, 30, 46];
or the dilemmas are presented in virtual reality rather than in text format [31, 32];
and perhaps many others yet to discover.

In all these cases, the bearing of these circumstances on UR also depends on the individual sensitivities. But given one same sensitivity level, their presence or absence weighs on the balance. Dilemmas where they are absent are easy for subjects with a strong deontological sensitivity and receive a swift deontological response. Dilemmas where one, many, or all of these circumstances are present are easy for subjects with strong utilitarian sensitivity. In some cases, they could be so easy that utilitarian responses would be intuitively issued. In the model by Baron and collaborators, circumstances of this type seem to play no role.

A paper by Krajbich and collaborators [33] explores a more suitable comparison than the comparison with semantic competence. The comparison is with public goods games (PGG). In such games, subjects are also torn between two sensitivities that oppose each other and are, when they conflict, exactly the converse of the other one: the selfish and the pro-social sensitivity. They often conflict, but not always, similar in this to the utilitarian and the deontological sensitivities. In the PGG, the difficulty refers to overcoming selfishness, which depends on objective features of the payoff structure. This is easily explained: If your contribution to the common fund generates for each group member, including yourself, a return only slightly below your contribution, it is easy to overcome the selfish inclination to contribute nothing to the public good. If on the contrary, it generates a return greatly below your contribution, it is not easy to overcome selfishness, because you risk losing virtually all your contribution if nobody else contributes [33]. People vary in the strength of their selfish and pro-social sensitivities, but this variance is always relative to those payoff structures. Krajbich et al. want to use this insight to criticize the dual-process model and favor a single process account. I believe this does not necessarily follow. Alternatively, you can argue that moral cognition depends essentially on emotional sensitivities. In particular, whether a given judgment or response to a moral dilemma or PGG is intuitive or controlled depends on the relative strength of the responder's opposed sensitivities.

A bewildering possibility is that some subjects could totally lack either the utilitarian or the deontological sensitivity. In these cases, subjects will give a response with no detection of conflict at all. Conflict-less responses can be labeled intuitive. Consider the percentage of UR to *footbridge*, which vary across studies roughly between 10% and 30%. Although consistently a minority, it is not an insignificant one. How do we interpret these participants? I see two possibilities: they feel the deontological intuition against the sacrifice and nonetheless decide that it is appropriate, or they feel no deontological intuition at all. The first case would correspond to the archetypal—though controversial—utilitarian subjects that Greene might have in mind, who out of conviction override their deontological intuitions. In the second case, however, it is hard to decide whether these participants, totally lacking a deontological sensitivity, have a moral sensitivity at all. Here several studies reporting positive correlations between UR and subclinical psychopathic tendencies become relevant. The correlations are small to moderate [34], and in all fairness, some studies have not found them [32], but in any case they might indicate that at least some subjects deliver UR score very low on empathy or high in clinical or subclinical psychopathy [24, 35–41], measured with psychometric questionnaires like the Levenson Self-Report Psychopathy Scale [42]. It is of course possible that participants lacking deontological intuitions are only a small minority within the group of up to 30% of participants that respond as utilitarians in *footbridge*. The rest are hard-core utilitarians, so to say, that override their deontological intuitions. For the sake of symmetry, one would suspect a similar situation for some deontological responses without conflict. They might reflect a cold-hearted rule following and a scant moral sensitivity [24]. Finding out if this is the case should be a goal for empirical research.

6 An Alternative Model of Moral Cognition

If we abandon the neat overlap between the deontological/utilitarian divide and the intuitive/reflexive divide, both Greene et al.'s particular dual-process model of moral cognition and Greene's normative conclusion should give way to an alternative version of the dual-process model and to a different normative conclusion. The alternative model contemplates both automatic utilitarian dispositions targeting group welfare and automatic deontological dispositions that partly conflict with them by protecting the individual against extreme group demands. When there is a conflict between utilitarian and deontological dispositions, the tension is real and cognitive control might take over (although we cannot assert with confidence that it always takes over). However, engagement of cognitive control does not necessarily lead to UR; deontological responses are also possible.

How should we picture the role of executive cognitive processes when they are engaged in tasks with moral dilemmas? In principle, cognitive control evaluates whether special circumstances speak in favor of UR or not. What kinds of circumstances are relevant? We already mentioned them above. Variables like a guilty or doomed victim, or the fact that the protagonist has stakes in the sacrifice (saves her own life), have a significant effect on the responses of participants relative to dilemmas where they are absent, like *footbridge* and *vitamins* [13, 28, 29, 43]. This increase has been confirmed with a battery that isolates the different contextual variables to different dilemmas, instead of including several in one (often the case in the items in Greene et al.'s battery and in most of its subsequent versions) and eliminating babies or children as victims [24]. The reasonable inference is that the additional circumstances (the doomed or guilty victim, or the selfish stakes in the sacrifice) are responsible for the increase, because these are the only elements that change from *footbridge* to, for example, *submarine*. In contrast, the judgment "saving five lives is better than saving one" remains constant. For this reason, if participants engage cognitive control in high-conflict dilemmas, it is probably to attend to these other variables and compute their effect on the decision. The increase in UR in the presence of these variables tells us that people pay special attention to them.

I shall now review experiments that provide evidence, sometimes indirectly, for the influence of each of these variables, beginning with doomed victims. Trémolière and Bonnefon [18] measured the UR as a function of the kill-save ratio and cognitive load. When the kill-save ratio is 1:5 cognitive, load interferes with the UR in *crying baby* and *captive soldier*. Participants under extreme load give significantly less UR than participants under light load. But when the ratio was 1:500, load did not interfere with UR in the same dilemmas. This suggests that when the kill-save ratio is not extraordinary, load interferes with processing the special circumstance of these dilemmas (doomed victim). When the kill-save ratio is extraordinary, it encourages all by itself and, intuitively [18], an increase of UR, making superfluous the controlled processing of other dilemma features. It remains to be investigated if extreme load would decrease the UR in dilemmas lacking special circumstances (like *footbridge*).

Other studies also suggest, indirectly, that participants use cognitive control to take the "doomed victim" feature into account. In an experiment designed to find evidence of the role of reflection and reasoning in moral judgment, Paxton et al. [23] tested participants with the Cognitive Reflection Test (CRT) [44] in two conditions—before and after responding to three high-conflict personal dilemmas—*footbridge*, *submarine*, and *crying baby*. Participants who responded to these dilemmas after the CRT showed a significant increase in utilitarian responses compared to participants who answered dilemmas before the CRT. Placing the CRT before the dilemmas primed participants to reflect when responding to them. But significantly, this effect was found only in *submarine* and in *crying baby*, and not in *footbridge* ([23], p. 168). They do not make much of this result, but the following explanation is plausible. When participants were primed, their reflections did not particularly target the utilitarian calculus that five is better than one (the only relevant factor present in *footbridge* and for which perhaps not much reflection is needed) but the fact that the person to be sacrificed would die anyway, a circumstance affecting *submarine* and *crying baby*, but not *footbridge*. This fact, when present, can reasonably be taken to shift the balance in favor of UR. A study by Moore et al. [13] targeted this variable directly. They investigated the effect of working memory capacity in utilitarian responses, controlling for factors like benefiting from the sacrifice or not, killing a person doomed to die or not, or killing as a means vs. killing as a side effect and without personal force. They found that participants with higher working memory scores gave significantly more UR than those with lower scores when the killing is personal and the victim is doomed to die anyway. They found no effect of working memory capacity in personal dilemmas like *footbridge*. This suggests that working memory is not engaged to compute the mere utilitarian benefit, but rather the fact that the victim is doomed to die.

Another circumstance that shifts the balance in favor of UR was disclosed in the pioneering experiments of Greene et al. They demonstrate that in impersonal dilemmas, where the loss of life results as a side effect and without exerting muscular force, most people normally condone the loss of life. Greene has argued that both features of impersonal killing are unjustified automatic settings of our moral minds. He claims, for example, that no moral difference exists between an intentional killing and one that, though not intended, is foreseen with certainty ([5], pp. 223–225). I beg to differ. I think this shows precisely how utilitarian intuitions conflicting with deontology effectively shape some of our decisions when aided by special circumstances. In this case, the special circumstance is the lack of intention to harm. To give a real-life example of a case like this one, recall Mackie's common sense explanation of why societies and states condone the loss of life statistically predicted as a side effect of motor vehicle transportation. The reason is, Mackie conjectures, that the benefits of getting faster to destination outweigh the disadvantages of lives lost, or so most of us think, consciously or not. These losses are statistically foreseen side effects, but not something that we want or intend ([45], p. 195). I think this example also brings vividly to awareness how some of our actual practices reveal a utilitarian influence that we could actually feel, after reflection, as deontologically suspect. Apparently, we humans tend to be influenced by utilitarian considerations in our moral practices and also in our judgments. Similarly, some circumstances can legitimate constraints on individual freedom—consider, for example, the measures that state and society could implement to prevent local population explosions. Those measures usually invade the (deontological) rights of the individual for the good of the group (the nation).

The other component of impersonal killing that favors UR, namely, the lack of muscular force, is certainly bizarre. Greene has insisted, correctly, that it is morally irrelevant. It could be just a hardwired and inaccurate proxy for unintended harm, functional in ancestral times, but not today. Participants in experiments do not confuse the exertion of muscular force with intention to harm, as shown by the *obstacle collide* scenario, a variant of *footbridge* where the death of the victim is caused with exertion of force but not as a means to save the five workmen ([5], pp. 218–202). But in contrast, participants seem to take the absence of muscular force for absence of intention to harm. When the victim is treated intentionally as a means to save others, but without the exertion of muscular force (Mikhail's *drop man* scenario), UR increases from 10% to 62% ([46], p. 149). The lack of muscular force increases the disposition to condone the loss of life in *drop man*, in spite of the fact that intention to harm is present in that scenario. Quite a lot of people, therefore, get things wrong and the reason seems to lie in an intuitive reaction, triggered by the automatic settings of our minds [5]. It remains to be investigated, however, whether participants scoring high in cognitive reflection, or induced to reflect before responding, are able to override its influence.

Of the variables that increase UR, one of the strangest was disclosed by two experiments that confronted subjects with virtual reality versions of personal and impersonal dilemmas. Though this mode of presentation increases emotional arousal (measured physiologically in both studies), results show, against all expectations, that it also increases UR, both in impersonal [31] and personal dilemmas [32]. In both cases, the authors explain this result with Cushman's version [47] of the dual-process model, where the processes in question concern the value of actions vs. the value of outcomes. It so happens that the virtual reality mode of presentation gives the five deaths resulting from inaction a stronger negative value than the action of killing one person. This poses an interesting challenge to interpretation, but I shall not attempt one here.

Other variables in Greene et al.'s original battery increase UR. When the victim is guilty, it is not excluded that at least some—and perhaps most—participants deliver an "intuitive" UR, as noted in the discussion of *preventing the spread* in Sect. 4 above, although I also noted that in this case it is actually difficult to distinguish it from an intuitive deontological response. It could well be a case of congruence between utilitarian and deontological intuitions, at least for some, or perhaps most participants. Another well-documented feature increasing UR is when agents benefit from the sacrifice: the fact that she is going to save her own life, not just the lives of several others—which, note, is not the case of *footbridge*—produces an increase in UR [13, 24]. Here it is plausible to postulate an automatic selfish response. Moore et al. [13] found that participants with greater working memory capacity do not give more UR in selfish dilemmas than participants with lower capacity. But Rand et al. [48] have found that pro-social responses, rather than selfish ones, are actually intuitive in the public goods game. How can we reconcile both results? Following our interpretation of Krajbich et al. [33] and the general gist of our preferred dual-process model, deontological or utilitarian responses are not per se intuitive or reflective but are one or the other depending on the particular individual sensitivities and the objective circumstances whose presence/absence speaks to those sensitivities.

We can apply this idea to all the circumstances that research has shown to increase UR. We could test each circumstance separately with the method of cognitive load, as in some papers reviewed above [15, 18, 21]. If we find that some of these circumstances increase UR independently of extreme load, this is evidence that they influence most individuals independently of working memory. If, however, the increase of UR is affected negatively by extreme load, this is evidence that most subjects need to compute them into the decision. In between, there is more individual variability, and we should not forget the possibility of cultural variability as well.

If this is how we should proceed to discern intuitive from controlled processing in moral cognition, this should also transfer to the design of experiments for collecting fMRI data. The procedure must be similar in both cases. Just as we test case by case the effect of load on the circumstances that increase UR, we should test case by case to observe how the fMRI data relate to the findings obtained from the load experiments. In this way, we can detect the instances where cognitive control attends to and ponders the circumstances that potentially justify a violation of the deontological rule. And if despite attending and pondering, the response is deontological, this should be taken as evidence that deontological responses can also arise from cognitive control.

7 The Normative Conclusion

The alternative version of the dual-process theory of moral cognition presents utilitarianism and deontology as two different moral intuitions hardwired by natural selection into our brains/minds. They are partially different and equally fundamental. This means that we are designed with a moral ambivalence. This is no surprise, for by now we know that some degree of imperfection indicates the hand of natural selection. Depending on the circumstances, some degree of interference, for the good of the group, with otherwise legitimate individual freedom will be condoned in a given society or culture. Taxes may come to mind as an example, but since taxes are so familiar to us all, no one except political philosophers would say that they violate deontological freedoms. A less familiar but not altogether distant example is the punishment that states implement to control local population explosion for the good of the group. This is a better candidate for (deontologically) illegitimate state control. Inevitably, the solutions to moral ambivalence will vary across cultural, geographical, and historical divides [49]. Thus, fundamental disagreement arises between societies and cultures, as it often arises within them.

What does our normative conclusion consist in? Greene anchored his normative conclusion in a theory over the standards of rational moral discourse. Rational moral discourse must be deliberative and argumentative in pursuit of the common good. Following singular intuitions cannot be the right track. I agree that this consideration is important and that it favors whatever moral view satisfies it. But the neurocognitive data collected in experimentation might still tell us that deontological responses satisfy it as well. I believe that we ought to recognize that deontological and utilitarian intuitions are often the boundaries within which our moral deliberations move freely and that any theory that would discount deontological principles and claims as nonrational would fail to satisfy the standards of deliberation. Counting heads is important, but several other things are important as well. The freedoms of individuals are important and so are the circumstances favoring head-count decisions in cases of conflict. But these circumstances are not written in the stars. The tension between utilitarian and deontological values is real and we have no innate guidance to resolve it. Deliberation remains a requirement for moral decisions, but deliberation trades in those two values (and possibly others). Different solutions arise in different times and places and in different heads and hearts. Normatively, there is no superiority of utilitarianism over deontology or the contrary, and no resolution of their conflict has any context-independent normative authority over any other. Thomas Nagel, not bothering to mention imperfect evolutionary design, has referred to this view as the “fragmentation of value” [50]. If my interpretation of the available neurocognitive data is correct, we are invited to embrace the “fragmentation of value,” rather than full-blown utilitarian morality.

Acknowledgments I am grateful to Josh Greene for sharing the data of the 2004 *Neuron* paper, to an anonymous reviewer for helpful comments, and to Wim De Neys for sharing work in press during the writings of this paper.

References

1. Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. Science. 2001;293:2105–8. https://doi.org/10.1126/science.1062872.
2. Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. The neural bases of cognitive conflict and control in moral judgment. Neuron. 2004;44:389–400. https://doi.org/10.1016/j.neuron.2004.09.027.
3. Tomasello M, Carpenter M. Shared intentionality. Dev Sci. 2007;10(1):121–5. https://doi.org/10.1111/j.1467-7687.2007.00573.x.
4. Greene J. From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? Nat Rev Neurosci. 2003;4:847–50.
5. Greene J. Moral tribes: emotion, reason and the gap between us and them. New York: Penguin Press; 2013.
6. Kahneman D. Thinking, fast and slow. New York: Farrar, Strauss and Giroux; 2011.
7. Evans JBT. Dual-processing accounts of reasoning, judgment, and social cognition. Annu Rev Psychol. 2008;59:255–78.
8. Diamond A. Executive functions. Annu Rev Psychol. 2013;64:135–68.
9. Greene JD, Morelli SA, Lowenberg K, Nystrom LE, Cohen JD. Cognitive load selectively interferes with utilitarian moral judgment. Cognition. 2008;107:1144–54. https://doi.org/10.1016/j.cognition.2007.11.004.
10. Kant I. Critique of practical reason. Indianapolis: Hackett; 2002 [1788].
11. McGuire J, Langdon R, Coltheart M, Mackenzie C. A reanalysis of the personal/impersonal distinction in moral psychology research. J Exp Soc Psychol. 2009;45(3):581–4. https://doi.org/10.1016/j.jesp.2009.01.002.
12. Koop GJ. An assessment of the temporal dynamics of moral decisions. Judgm Decis Mak. 2013;8(5):527–39.
13. Moore AB, Clark BA, Kane MJ. Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. Psychol Sci. 2008;19:549–57. https://doi.org/10.1111/j.1467- 9280.2008.02122.x.
14. Greene J. Dual-process morality and the personal/impersonal distinction: a reply to McGuire, Langdon, Coltheart, and Mackenzie. J Exp Soc Psychol. 2009;45:581–4. https://doi.org/10.1016/j.jesp.2009.01.003.
15. Bialek M, De Neys W. Dual processes and moral conflict: evidence for deontological reasoners' intuitive utilitarian sensitivity. Judgm Decis Mak. 2017;12(2):148–67.
16. Dubljević V, Racine E. The ADC of moral judgment: opening the black box of moral intuitions with heuristics about agents, deeds, and consequences. AJOB Neurosci. 2014;5:3–20.
17. Kahane G, Wiech K, Shackel N, Farias M, Savulescu J, Tracey I. The neural basis of intuitive and counterintuitive moral judgment. Soc Cogn Affect Neurosci. 2012;7:393–402. https://doi.org/10.1093/scan/nsr005.
18. Trémolière B, Bonnefon JF. Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. Pers Soc Psychol Bull. 2014;40:923–30.
19. De Neys W. Bias and conflict: a case for logical intuitions. Perspect Psychol Sci. 2012;7:28–3.
20. Bago B, De Neys W. Fast logic? Examining the time course assumption of dual process theory. Cognition. 2017;158:90–109.
21. Białek M, De Neys W. Conflict detection during moral decision-making: evidence for deontic reasoners' utilitarian sensitivity. J Cogn Psychol. 2016;28(5):631–9. https://doi.org/10.1080/20445911.2016.1156118.

22. Nichols S, Mallon R. Moral dilemmas and moral rules. Cognition. 2006;100(3):530–42.
23. Paxton JM, Ungar L, Greene J. Reflection and reasoning in moral judgment. Cogn Sci. 2012;36:163–77.
24. Rosas A, Viciana H, Caviedes E, Arciniegas A. Hot utilitarianism and cold deontology: insights from a response-patterns approach to sacrificial and real world dilemmas. Submitted.
25. Baron J, Gürçay B, Moore AB, Starcke K. Use of a Rasch model to predict response times to utilitarian moral dilemmas. Synthese. 2012;189(S1):107–17. https://doi.org/10.1007/s11229-012-0121-z.
26. Baron J, Gürçay B. A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. Mem Cogn. 2016;45:566. https://doi.org/10.3758/s13421-016-0686-8.
27. Conway P, Gawronski B. Deontological and utilitarian inclinations in moral decision making: a process dissociation approach. J Pers Soc Psychol. 2013;104(2):216–35. https://doi.org/10.1037/a0031021.
28. Huebner B, Hauser MD, Pettit P. How the source, inevitability and means of bringing about harm interact in folk-moral judgments. Mind Lang. 2011;26:210–33. https://doi.org/10.1111/j.1468-0017.2011.01416.x.
29. Rosas A, Koenigs M. Beyond 'utilitarianism': maximizing the clinical impact of moral judgment research. Soc Neurosci. 2014;9:661–7. https://doi.org/10.1080/17470919.2014.937506.
30. Greene JD, Cushman FA, Stewart LE, Lowenberg K, Nystrom LE, Cohen JD. Pushing moral buttons: the interaction between personal force and intention in moral judgment. Cognition. 2009;111:364–71. https://doi.org/10.1016/j.cognition.2009.02.001.
31. Patil I, Cogoni C, Zangrando N, Chittaro L, Silani G. Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. Soc Neurosci. 2014;9(1):94–107. https://doi.org/10.1080/17470919.2013.870091.
32. Francis KB, Howard C, Howard IS, Gummerum M, Ganis G, Anderson G, Terbeck S. Virtual morality: transitioning from moral judgment to moral action? PLoS One. 2016;11(10):e0164374. https://doi.org/10.1371/journal.pone.0164374.
33. Krajbich I, Bartling B, Hare T, Fehr E. Rethinking fast and slow based on a critique of reaction-time reverse inference. Nat Commun. 2015;6:7455. https://doi.org/10.1038/ncomms8455.
34. Kahane G, Everett JAC, Earp BD, Farias M, Savulescu J. 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. Cognition. 2015;134:193–209. https://doi.org/10.1016/j. cognition. 2014.10.005.
35. Bartels DM, Pizarro D. The mismeasure of morals: antisocial personality traits predict utilitarian responses to moral dilemmas. Cognition. 2011;121:154–61. https://doi.org/10.1016/j.cognition.2011.05.010.
36. Djeriouat H, Trémolière B. The dark triad of personality and utilitarian moral judgment: the mediating role of honesty/humility and harm/care. Personal Individ Differ. 2014;67:11–6. https://doi.org/10.1016/j.paid.2013.12.026.
37. Duke AA, Bègue L. The drunk utilitarian: blood alcohol concentration predicts utilitarian responses in moral dilemmas. Cognition. 2015;134:121–7. https://doi.org/10.1016/j.cognition.2014.09.006.
38. Glenn AL, Koleva S, Iyer R, Graham J, Ditto PH. Moral identity in psychopathy. Judgm Decis Mak. 2010;5(7):497–505.
39. Gleichgerrcht E, Young L. Low levels of empathic concern predict utilitarian moral judgment. PLoS One. 2013;8(4):e60418. https://doi.org/10.1371/journal.pone.0060418.
40. Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A. Damage to the prefrontal cortex increases utilitarian moral judgments. Nature. 2007;446:908–11. https://doi.org/10.1038/nature0563.
41. Patil I. Trait psychopathy and utilitarian moral judgment: the mediating role of action aversion. J. Cogn Psychol. 2015;27(3):349–66. https://doi.org/10.1080/20445911.2015.1004334.
42. Levenson MR, Kiehl KA, Fitzpatrick CM. Assessing psychopathic attributes in a noninstitutionalized population. J Pers Soc Psychol. 1995;68(1):151–8.

43. Christensen JF, Flexas A, Calabrese M, Gut NK, Gomila A. Moral judgment reloaded: a moral dilemma validation study. Front Psychol. 2014;5:607. https://doi.org/10.3389/fpsyg.2014.00607.
44. Frederick S. Cognitive reflection and decision making. J Econ Perspect. 2005;19:25–42.
45. Mackie JL. Ethics: inventing right and wrong. London: Penguin Group; 1977.
46. Mikhail J. Universal moral grammar: theory, evidence and the future. Trends Cogn Sci. 2007;11(4):143–52.
47. Cushman F. Action, outcome, and value: a dual-system framework for morality. Personal Soc Psychol Rev. 2013;17(3):273–92. https://doi.org/10.1177/1088868313495594.
48. Rand DG, Greene JD, Nowak MA. Spontaneous giving and calculated greed. Nature. 2012;489:427–30.
49. Wong D. Natural moralities. A defense of pluralistic relativism. Oxford: Oxford University Press; 2006.
50. Nagel T. The fragmentation of value. In: Nagel T. Mortal questions. Cambridge: Cambridge University Press; 1979. p. 128–41.