Unequal Vividness and Double Effect

NEIL SINHABABU

National University of Singapore

I argue that the Doctrine of Double Effect is accepted because of unreliable processes of belief-formation, making it unacceptably likely to be mistaken. We accept the doctrine because we more vividly imagine intended consequences of our actions than merely foreseen ones, making our aversions to the intended harms more violent, and making us judge that producing the intended harms is morally worse. This explanation fits psychological evidence from Schnall and others, and recent neuroscientific research from Greene, Klein, Kahane and Schaich Borg. It explains Mikhail and Hauser's 'universal moral grammar' and an interesting phenomenon about Double Effect cases noted by Bennett. When unequally vivid representations determine our decisions, we typically misjudge the merits of our options and make mistakes. So if Double Effect is a product of unequal vividness, it is likely to be mistaken. This argument, I claim, fits Berker's specifications for good empirically grounded arguments in ethics.

Joshua Greene and Peter Singer have recently argued against deontological principles on neuroscientific grounds, claiming that the emotional processes producing these principles lead us into error.[1] I concur with Selim Berker's assessment of their arguments:

> either attempts to derive normative implications from these neuroscientific results rely on a shoddy inference, or they appeal to substantive normative intuitions (usually about what sorts of features are or are not morally relevant) that render the neuroscientific results irrelevant to the overall argument.[2]

Recent work from Colin Klein and Guy Kahane has cast some of the neuroscientific presuppositions of Singer and Greene's arguments into doubt.[3]

[1] Joshua Greene, 'The Secret Joke of Kant's Soul', *Moral Psychology*, vol. 3: *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, ed. Walter Sinnott-Armstrong (Cambridge, Mass., 2008), pp. 35–79; Peter Singer, 'Ethics and Intuitions', *The Journal of Ethics* 9 (2005), pp. 331–52. Much of the empirical data supporting these positions appears in Joshua Greene et al., 'An fMRI Investigation of Emotional Engagement in Moral Judgment', *Science* 293 (2001), pp. 2105–8; Joshua Greene et al., 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment', *Neuron* 44 (2004), pp. 389–400; and Joshua Greene et al., 'Cognitive Load Selectively Interferes with Utilitarian Moral Judgment', *Cognition* 107 (2008), pp. 1144–54.

[2] Selim Berker, 'The Normative Insignificance of Neuroscience', *Philosophy and Public Affairs* 37 (2009), pp. 293–329, at 294.

[3] Guy Kahane et al., 'The Neural Basis of Intuitive and Counterintuitive Moral Judgment', *Social Cognitive and Affective Neuroscience Advance Access* (2011), pp. 393–402; Colin Klein, 'The Dual Track Theory of Moral Decision-Making: A Critique of the Neuroimaging Evidence', *Neuroethics* (2011), pp. 143–62.

© Cambridge University Press 2013
Utilitas Vol. 25, No. 3, September 2013
doi:10.1017/S0953820812000362

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

I hope to advance this debate with an argument that uses recent empirical results against the Doctrine of Double Effect as Greene and Singer wish to, but which avoids the problems Berker notes, and which better fits the data. It runs as follows:

> [Unequal Vividness Explanation] Double Effect is accepted because of how unequally vivid representations of actions' intended and merely foreseen consequences affect our desires.
> [Unreliability Claim] When the effects of unequally vivid representations upon our desires are decisive in causing our judgements about what to do, we are usually mistaken.
> [Conclusion] In accepting Double Effect, we are likely to be mistaken.

First, I'll lay out the thesis that I take all defenders of Double Effect to be committed to – that harms intended are morally worse to produce than equal harms merely foreseen. Then I'll present the unequal vividness explanation, which explains the appeal of Double Effect in terms of the difference between how we imagine the intended consequences of our actions and the consequences we merely foresee, its effects on the violence of our passions, and the effects of these passions on moral judgement. I'll support the unequal vividness explanation by showing how recent neuroscientific results support it, how it simplifies theories that posit a universal moral grammar, and how it accounts for an odd feature of Double Effect cases noted by Jonathan Bennett. Then I'll argue that the unequal vividness explanation gives us reason to think that Double Effect is mistaken. When unequally vivid representations determine our decisions, we usually decide wrongly. Irrational choices where we sacrifice greater distant goods for lesser nearby goods are typical examples. I'll consider three ways of objecting to the argument – denying the unequal vividness explanation, straightforwardly denying the unreliability claim, and arguing that the unreliability claim doesn't apply to Double Effect. Finally, I'll argue that my argument fits Berker's specifications for a successful empirically grounded argument, and defend a larger role for empirical methods in ethics than he allows.

WHAT IS DOUBLE EFFECT?

The Doctrine of Double Effect may be best introduced with examples. First, the example of the strategic bomber and the terror bomber, as presented by Alison McIntyre:[4]

> The terror bomber aims to bring about civilian deaths in order to weaken the resolve of the enemy: when his bombs kill civilians this is a consequence that he intends. The strategic bomber aims at military targets while foreseeing that

[4] Alison McIntyre, 'Doing Away With Double Effect', *Ethics* 111 (2001), pp. 219–55.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

bombing such targets will cause civilian deaths. When his bombs kill civilians this is a foreseen but unintended consequence of his actions. Even if it is equally certain that the two bombers will cause the same number of civilian deaths, terror bombing is impermissible, while strategic bombing is permissible.[5]

Second, the classic trolley examples:

It would be wrong to throw someone into the path of a runaway trolley in order to stop it and keep it from hitting five people on the track ahead; that would involve intending harm to the one as a means of saving the five. But it would be permissible to divert a runaway trolley onto a track holding one and away from a track holding five: in that case one foresees but does not intend the death of the one as a side effect of saving the five.[6]

Both examples compare two ways for an action to produce the same overall harms and benefits. In cases of terror bombing and pushing someone (as the example is usually presented, a fat man) in the trolley's path to block it, the agents intend harm as a means to produce some good end. In the strategic bomber and diverting cases, the agents produce a good end and don't use anyone as a means. They merely foresee that the same harms as in the terror bomber and throwing cases will then result. This is why the actions in the terror bomber and throwing cases are impermissible while those in the strategic bomber and diverting cases are permissible.

McIntyre identifies two claims made by proponents of Double Effect:

(1) it is sometimes permissible to bring about a harm as a merely foreseen side effect of an action aimed at some good end, even though it would have been impermissible to bring about the same harm as a means to that end, and (2) this is so because of the moral significance of the distinction between intending and foreseeing a harmful consequence of one's own agency.[7]

The distinction between intended and foreseen harms is at the heart of most formulations of Double Effect. On Neil Delaney's formulation,

in some cases it is worse to directly intend a bad thing as a means to a good end than it is merely to foresee that an equally bad thing will come about as a result of actions which are in themselves morally neutral at worst.[8]

[5] McIntyre, 'Doing Away With Double Effect', p. 219.

[6] McIntyre, 'Doing Away With Double Effect', p. 220.

[7] McIntyre, 'Doing Away With Double Effect', p. 219. McIntyre argues that Double Effect examples can't be captured under any simple principle. They provide a 'gallery of miscellaneous objections to simple forms of direct consequentialism that can be expressed, with more or less strain, using the distinction between intended and merely foreseen consequences. They are tied together by nothing more penetrating than the claim that the distinction between what an agent foresees and what an agent intends sometimes matters, and matters a great deal, to moral evaluation' ('Doing Away With Double Effect', p. 255).

[8] Neil Delaney, 'A Note on Intention and the Doctrine of Double Effect', *Philosophical Studies* 134 (2007), pp. 103–10, at 103.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

While some formulations aren't explicitly put in terms of the intended/foreseen distinction, they're naturally characterized that way. For example, Frances Kamm writes that 'it is permissible to do what is not in itself bad (or omit an act) though this has a bad side effect, if the good we seek to achieve is greater than that bad'.[9] Here it's natural to take what we 'seek to achieve' as what we intend, and its 'side effects' as what we merely foresee.

Henceforth I'll take Double Effect to be the view that sometimes harms intended are morally worse to produce than equal harms merely foreseen.[10] This formulation most closely follows Delaney. It also encompasses Kamm and McIntyre's formulations, as it's worse to do the impermissible than the permissible. Defeating this weak formulation of Double Effect will defeat any existing version of the principle.

THE UNEQUAL VIVIDNESS EXPLANATION OF DOUBLE EFFECT

Now I'll explain why Double Effect has such appeal to us. As we deliberate, we vividly imagine what we intend. We don't imagine what we merely foresee with the same vividness. The more vividly something bad is represented in our senses or our imagination, the more emotion our desires against it produce. So we feel stronger negative emotions against intended harms than merely foreseen harms. Since stronger negative emotions as we consider something make us see it as morally worse, we regard the intended harms as morally worse. When we use Double Effect to evaluate others' actions, all these factors affect how we understand their actions to look from their point of view. I'll lay out this picture in more detail and support it with evidence from recent psychological research.

We imagine intended consequences of our actions more vividly than merely foreseen ones. For example, when I decide to drive to the store, I imagine my car moving towards the store more vividly than I imagine my car making noise. This is because I intend to get there, while I merely foresee the noise. But if part of my intention is noise-related – perhaps I also intend to listen to my engine to test whether it's still making a strange sound that suggests I need repairs – I'll imagine the noise vividly as well.

The phenomenology of action would be very different if we imagined the merely foreseen as vividly as the intended. The number of things

[9] Frances Kamm, 'The Doctrine of Double Effect', *Journal of Medicine and Philosophy* 16 (1991), pp. 571–85, at 571–2.

[10] Here I must take 'harms intended' to include harms close to what we intend. The 'straightforward view' of intention in the section on Bennett's phenomenon will characterize them as not really intended.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

we expect to happen in the future as we follow some course of action is large, and the number of things we intend forms only a small proportion of them. What we intend is particularly salient to us. When my car is stopped and I intend to make a U-turn, I foresee all sorts of consequences of my actions, for example that my car will make noise, that the cup in the cup holder will move along with the car, that I'll feel the swerving, and so on. (I'd feel surprise if these things didn't happen.) But successfully causing the car to go the other way – the consequence I intend – is most salient to me as I get ready for the U-turn. None of these things might be more salient than the others to an idle passenger who merely foresees what is going to happen. But drivers' intentions cause them to focus on the car's direction.

Representing something more vividly increases the strength of our emotions about it, as Hume describes in distinguishing calm and violent passions.[11] The calm passions, 'tho' they be real passions, produce little emotion in the mind, and are more known by their effects than by the immediate feeling or sensation'. Violent passions, on the other hand, are experienced more robustly. Hume remarks that: 'When I am immediately threatened with any grievous ill, my fears, apprehensions, and aversions rise to a great height, and produce a sensible emotion.' This doesn't mean that each of our passions is fixed in its calmness or violence. By varying the situation of the object we can 'change the calm and violent passions into each other'.[12] A calm passion becomes violent when the object of the passion comes closer to the agent. As Hume says, 'The same good, when near, will cause a violent passion, which, when remote, produces only a calm one.' The way nearby goods increase the violence of passions is explained in turn by their greater vividness:

> There is an easy reason, why every thing contiguous to us, either in space or time, shou'd be conceiv'd with a peculiar force and vivacity, and excel every other object, in its influence on the imagination. Ourself is intimately present to us, and whatever is related to self must partake of that quality. But where an object is so far remov'd as to have lost the advantage of this relation ... its idea becomes still fainter and more obscure.[13]

Hume's view explains why we discount future goods in favour of more immediate satisfactions. This happens even when the probability of attaining the future goods if we choose them is as high as the probability of attaining immediate satisfactions if we choose them. The future goods are represented less vividly and generate less violent passions, while the immediate goods are represented more vividly and

[11] David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge and P. H. Nidditch (Oxford, 1978), 2.3.3.

[12] Hume, *Treatise*, 2.3.4.

[13] Hume, *Treatise*, 2.2.7.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

generate more violent passions. This leads the nearer goods to look more appealing, moving us to choose them.

Recent empirically oriented discussions of motivational phenomena have incorporated Hume's insight, noting how representations the agent associates with objects of desire increase the violence of that desire. According to neuroscientist Kent Berridge:

> One feature of incentive salience is to endow reward-related cues (in experiments these are Pavlovian conditioned stimuli or Css) with an ability to trigger powerful peaks of 'wanting' for their own associated reward. For example, the scent of food may suddenly make you ravenous as lunchtime approaches even if you were not feeling particularly hungry moments before that cue occurred.[14]

I invoke this property of desire to explain various motivational phenomena:

> when an agent is presented with vivid images she associates with a state of affairs she desires, either in imagination or by her senses, that will strengthen the desire's causal powers. The desire's phenomenal effects increase greatly, and its motivational powers increase substantially as well.[15]

Phenomenal and motivational effects are the two kinds of effects that Hume has discussed – the 'sensible emotion' desire produces, and its influence over our actions. This property of desire will cause intended harms, which we imagine more vividly, to have a greater impact on our feelings and behaviour than foreseen harms, which we imagine less vividly.

Our emotions as we consider something influence our moral judgements about it. This thesis has been borne out by a series of recent experiments in which Simone Schnall and colleagues manipulated subjects' emotions so as to generate emotions of mild disgust, causing negative moral evaluations of things entirely unrelated to the causes of the disgust. In one series of experiments, some participants filled in questionnaires about the rightness or wrongness of various actions in rooms with a disgusting smell, while other participants filled in the questionnaires in rooms that had no smell at all.[16] The subjects who were exposed to the disgusting smell rated the actions as being more wrong. Another experiment had some participants filling in the questionnaires in a very tidy room and being told that they should

[14] Kent Berridge, 'Wanting and Liking: Observations from the Neuroscience and Psychology Laboratory', *Inquiry*, 52 (2009), pp. 378–98, at 380.

[15] Neil Sinhababu, 'The Humean Theory of Motivation Reformulated and Defended', *Philosophical Review* 118 (2009), pp. 465–500, at 471.

[16] Simone Schnall et al., 'Disgust as Embodied Moral Judgment', *Personality and Social Psychology Bulletin* 34 (2008), pp. 1096–1109.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

wash their hands first.[17] These participants took the actions described in the survey to be less wrong than participants who were not sent to wash their hands and took the survey in a messy room. These results demonstrate how negative emotional states darken our negative moral evaluations, whether by making them more negative or more vehement. While they aren't Double Effect cases and don't involve killing, they illustrate the principle that negative emotions as we consider something generate negative moral judgements. This principle is borne out by other empirical findings. The inability to feel emotions like guilt and sadness is implicated in psychopaths' inability to judge that particular actions would be wrong.[18] If negative emotion is present as we consider something, negative moral judgement follows, while the absence of negative emotion leads to a lack of negative moral judgement.[19]

We now can explain the appeal of Double Effect. We imagine what we intend more vividly than what we merely foresee. Since vividly representing a passion's object increases the violence of the passion, the greater vividness of intended harms produces more violent passions against these harms. Being more violently struck by aversions to harms we intend than harms we merely foresee, we feel worse about producing the intended harms than the foreseen harms. Our more violent aversions to them move us to avoid causing them, and cause us to feel them as more severe moral violations. If the unequal vividness explanation is correct, our psychology produces moral judgement-affecting emotions in Double Effect cases much like the effects of a disgusting smell from Schnall's experiments.

This explanation of Double Effect is tailored to first-personal cases, where we are asked whether we would (for example) push a fat man over a bridge to block a trolley from killing five people. It can be expanded to fit third-personal cases where we evaluate others' actions if we add that considering others' actions from the first-person perspective is part of evaluating them. Then the phenomenological difference between harms intended and merely foreseen by the agent will be salient to us. There are many ways in which we take such

[17] Simone Schnall, Jennifer Benton and Sophie Harvey, 'With a Clean Conscience: Cleanliness Reduces the Severity of Moral Judgments', *Psychological Science* 19 (2008), pp. 1219–22. For more cases of how emotions affect moral judgement, see Jonathan Haidt, 'The Emotional Dog and the Rational Tail', *Psychological Review* 108 (2001), pp. 814–34.

[18] Hervey Cleckley, *The Mask of Sanity* (Augusta, 1988); R. J. R. Blair, 'A Cognitive Developmental Approach to Morality: Investigating the Psychopath', *Cognition* 57 (1995), pp. 1–29; Kent Kiehl, 'A Cognitive Neuroscience Perspective on Psychopathy: Evidence for Paralimbic System Dysfunction', *Psychiatry Research* 142 (2006), pp. 107–28.

[19] For more experimental findings along these lines, see Jesse Prinz, *The Emotional Construction of Morals* (Oxford, 2007).

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

a first-person view when evaluating others' actions. For example, we consider what agents believe, and withhold blame when their actions turn out badly just because of something they couldn't have known. If agents harm their friends or family members to whom we have no special relationship, we evaluate their actions by seeing the victims in the light of the agents' relationship to them, not ours. Given all the ways we assume a first-person perspective on those whose actions we judge, it isn't surprising that we appreciate the distinction between the intended and the merely foreseen from their perspective. Since the way we imagine harms in third-person Double Effect cases is mediated by an additional imaginative step (we have to imagine what someone imagined rather than imagining directly), we might expect the extra badness of the harms to be less vividly represented than in first-person cases. Perhaps the closest thing to a test of this claim in the literature is a comparison of first-person and third-person versions of the switching trolley case. While 90 per cent of respondents thought it would be permissible for someone else – 'John' – to divert the trolley, only 65 per cent thought it would be permissible to do so themselves.[20]

The unequal vividness explanation has several parts – the greater vividness of what we intend than what we merely foresee, the connection between vividness and the passions, the connection between the passions and moral judgement, and taking a first-personal perspective on someone else's action in third-personal cases. One might criticize this explanation for being excessively complex, and hope for some simpler account. Those with such misgivings should note that each part of the unequal vividness explanation can be confirmed independently, at least phenomenologically, and that some also have support from psychological research. That we imagine what we intend more vividly than what we merely foresee, that more vividly imagined things produce stronger emotions, that stronger emotions lead to stronger moral judgements, and that we morally evaluate others' actions from a first-personal perspective are all things that we can all discover by attending to our own experiences of planning, imagining, desiring and judging. I explain the appeal of Double Effect not by positing novel psychological structures specific to moral thought, but by noting how the natural operations of more basic psychological processes can't help but produce it.

Unequal vividness might explain some other ways in which simple consequentialist views can seem unappealing. For example, if we imagine the same amount of harm more vividly when it's concentrated in small numbers of people than when it's distributed among a

[20] Thomas Nadelhoffer and Adam Feltz, 'The Actor-Observer Bias and Moral Intuitions: Adding Fuel to Sinnott-Armstrong's Fire', *Neuroethics* 1 (2008), pp. 133–44.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

multitude, that would explain why utilitarians' aggregative principles can seem unappealing.[21] (I don't yet see that this hypothesis extends to deontological prohibitions against well-defined act-types like lying or stealing, unless these are themselves Double Effect cases.) I won't develop this hypothesis further here. Sympathetic readers will be able to develop it without further assistance, and unsympathetic readers will be sceptical enough about my attack on Double Effect that it'll be worthwhile for me to focus on that doctrine alone.

UNEQUAL VIVIDNESS AND NEUROSCIENCE

Now I'll consider how the unequal vividness explanation fits recent neuroscientific data about moral judgement. The research I'll discuss compares brain activity in cases where people make judgements about what it is right to do in a variety of moral dilemmas. Of interest are the contrasts between the psychological processes giving intuitive support to deontological and utilitarian conclusions. The existing data don't apply as well as one might hope to the unequal vividness explanation, as some of the dilemmas involve deontological prohibitions against lying or stealing, which unequal vividness may not explain. But some of the dilemmas involve Double Effect judgements, and others involve other deontological principles where the unequal vividness explanation is plausible. In view of the dialectical significance of this research, which provided the empirical backing for Singer and Greene's arguments, I should consider what it says about the unequal vividness explanation.

What sort of evidence would support the unequal vividness explanation? According to this explanation, the same basic process produces deontological and utilitarian judgements. In both cases, we imagine outcomes, have emotions about them and form judgements about what should be done. The difference is that when deontological principles like Double Effect are intuitive, we imagine the negative consequences of our actions more vividly, and feel more violent emotions about them. So we would expect more activation in regions associated with imagination and emotion in these cases. Since the unequal vividness explanation doesn't suggest more activity in any mental process when utilitarian judgements are intuitive, we would expect no regions to show greater activation in these cases.

Greene takes his brain-imaging research to support a dual-process model in which two different systems produce moral judgement –

[21] This also explains the appeal of choosing a normal life over a life of button-pushing to generate unspecified good results in the case at the end of James Lenman, 'Consequentialism and Cluelessness', *Philosophy and Public Affairs* 29 (2000), pp. 342–70.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

a deontological system governed by automatic emotional responses, and a utilitarian system engaging in more full-fledged reasoning.[22] He claims that the dorsolateral prefrontal cortex and the inferior parietal lobe, which are associated with reasoning, are more active in utilitarian judgements. Meanwhile, he claims that the precuneus (PC), the posterior cingulate cortex (PCC), the medial prefrontal cortex and the amygdala, which are associated with emotion, are more active in deontological judgements. Citing Nietzsche, he refers to Immanuel Kant's defence of the conclusions of the emotional deontological system with a theoretical approach that emphasized the moral significance of reason as 'the secret joke of Kant's soul'. While Greene's data support the unequal vividness explanation to some extent, with its emphasis on greater emotion in deontological judgements, the greater activity in cognitive regions for utilitarian judgements remains to be explained, and the stark division between the processes runs contrary to the relative unity of the unequal vividness explanation.

Colin Klein's criticisms of the neuroscientific presuppositions of Greene's dual-process model suggest that the processes producing the judgements are more similar and confirm the role of imagination. Perhaps Klein's starkest criticism is that Greene's 'claim that PC/PCC makes for a specifically emotional processing area is no longer sustainable'.[23] As he notes, the PC is 'most commonly associated with three types of task: visual imagery, successful episodic memory retrieval (both visual and non-visual), and self-referential processing'.[24] He continues to note that 'self-referential processing is also strongly implicated in taking a first-person perspective during tasks'.[25] He describes the PCC as similar to the PC, with both regions playing a role in 'self-reflexive processing and imagination'.[26] While this is trouble for any model presupposing that these regions are specifically emotional, it's a boon for the unequal vividness explanation, which invokes vivid imagination to explain how Double Effect intuitions arise and the adoption of a first-person perspective

[22] I've omitted mention of Greene's reaction time data, as it seems that Berker's arguments and Kahane's findings on that topic leave little reason to discuss them. See also Jonathan McGuire, Robyn Langdon, Max Coltheart, and Catriona Mackenzie, 'A Reanalysis of the Personal/Impersonal Distinction in Moral Psychology Research', *Journal of Experimental Social Psychology* 45 (2009), pp. 577–80.

[23] Klein, 'Dual Track Theory', p. 146.

[24] Klein, 'Dual Track Theory', p. 146, citing A. E. Cavanna and M. R. Trimble, 'The Precuneus: A Review of its Functional Anatomy and Behavioural Correlates', *Brain* 129 (2006), pp. 564–83.

[25] Klein, 'Dual Track Theory', p. 147, citing K. Vogeley and G. R. Fink, 'Neural Correlates of the First-Person-Perspective', *Trends in Cognitive Sciences* 7 (2003), pp. 38–42.

[26] Klein, 'Dual Track Theory', p. 147.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

to explain how these intuitions affect third-person moral judgements. Klein's other criticisms of Greene run along similar lines – pointing out that 'none of the cortical areas identified by Greene et al are functionally specific: each is active in a wide variety of both cognitive and emotional tasks'.[27] While this is a problem for a simple division of the brain into cognitive and emotional areas, it's what we'd expect if the unequal vividness explanation were correct. Vivid imagination is active in a wide variety of cognitive and emotional tasks.

Kahane and his collaborators provide an alternative explanation of the neural activation that Greene took as a sign of utilitarian judgements issuing from a special system. In the dilemmas Greene considered, the deontological response was intuitive while the utilitarian response was counterintuitive. Testing cases in which the utilitarian response was intuitive (UI cases) as well as cases in which the deontological response was intuitive (DI cases), Kahane found that this pattern of neural activation applied for counterintuitive judgements regardless of whether they were utilitarian or deontological. As he writes, 'Our findings thus suggest that even in the context of the extreme moral dilemmas previously studied, the neural activations associated with utilitarian judgments might be due to their counter-intuitiveness, not their content.'[28] Apart from the effects of intuitiveness, 'Utilitarian judgements did not exhibit any specific significant activations,' and neither did UI cases.[29] However, Kahane confirms Greene's discovery of greater PCC activation in deontological judgements. He finds more activity in the right temporoparietal junction (TPJ), which he says might 'reflect the central role of intention in determining permissibility in deontological ethics', or 'concern with one's own agency and its emotional significance'.[30] He similarly detects greater orbitofrontal cortex activation when deontological judgements are intuitive.[31] DI cases, similarly, showed greater activation than UI cases in the PCC and TPJ, as well as two regions of the prefrontal cortex.

Jana Schaich Borg and her collaborators provide evidence of greater emotional activation in Double Effect cases. As she tested Double

[27] Klein, 'Dual Track Theory', p. 143.

[28] Kahane et al., 'Neural Basis', p. 399.

[29] Kahane et al., 'Neural Basis', p. 399. One shortcoming of Kahane's experiments is that his DI cases and UI cases are entirely different, with different sets of harms. This is unlike McIntyre's trolley and bomber examples, which involve similar situations and pair the same harms against each other – 1 dead versus 5 in both trolley cases, and the same number of deaths in each bomber case. Kahane thus doesn't control for the possibility that one set of harms arouses more emotion in subjects than the other. Furthermore, five UI cases involve lying while no DI cases do.

[30] Kahane et al., 'Neural Basis', p. 401.

[31] Kahane et al., 'Neural Basis', p. 396.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

Effect specifically, this evidence is especially helpful in evaluating the unequal vividness explanation. She writes that 'Compared to moral scenarios involving only unintentional harm, moral scenarios involving intentional harm elicit more activity in areas associated with emotion (orbitofrontal cortex and temporal pole).'[32] Schaich Borg also tested cases of intentional harm both in plain language and in more vivid language. The more vivid descriptions raised the percentage of subjects saying it was wrong to act from 63 per cent to 88 per cent, and lowered the percentage of subjects saying they would act from 40 per cent to 13 per cent.[33] This supports the claim that more vividly represented harms make us judge actions to be wrong and make us less willing to act.

What's the upshot of these empirical results? There seems to be a difference between the psychological processes giving rise to utilitarian and deontological intuitions, though not as stark as the one posited by Greene. Areas responsible for imaginative representation and emotion seem to be more active when we make deontological judgements or consider cases where deontological judgements are intuitive. No areas are especially activated in cases where utilitarian judgements are intuitive. While the examples aren't all optimal for testing it, the results fit the unequal vividness explanation fairly well.

EXPLAINING THE UNIVERSAL MORAL GRAMMAR

Now I'll argue that unequal vividness accounts for the universality of Double Effect intuitions, providing an elegant explanation of what Marc Hauser and John Mikhail call the 'universal moral grammar'. Hauser and his collaborators collected a great deal of data confirming the intuitiveness of Double Effect among a broad range of different populations. Thirty thousand internet surveys of people in 120 countries showed that a very wide variety of demographic subgroups agreed that it was permissible to switch the track so that the trolley would kill one instead of five, and also that it was impermissible to throw the fat man over the bridge. The same results were found quite robustly among all demographic subsets for which they had a sufficiently large sample size – educational levels from elementary school to the doctoral level, all age groups over 10 years, both genders,

[32] Jana Schaich Borg et al., 'Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation', *Journal of Cognitive Neuroscience* 18 (2006), pp. 803–17, at 803. Schaich Borg also found less activity in the right angular gyrus and superior frontal gyrus in Double Effect cases, which she associates with cognition. She suggests that this results from subjects being dumbfounded by their intuitive judgements (p. 815).

[33] Borg et al., 'Consequences, Action, and Intention', p. 809.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

and all nations except Ireland (which was only a sample of 16). Overall, 89 per cent of people thought it was permissible to switch the track, but only 11 per cent thought it was permissible to push the fat man.[34] Mikhail and his collaborators found similar results in a variety of populations including children ages 8–12.[35]

To explain the robustness of these results along such a broad spectrum of different demographic groups, Hauser and Mikhail posit a 'universal moral grammar' constituted by rules about the rightness and wrongness of various actions, separated by features like their causal structure – what is intended as a means to what, for example, and what is merely foreseen. The analogy to Chomsky's views in linguistics is intended – just as the rules of universal grammar constrain the syntactic structures of possible languages while allowing a huge variety of languages to flourish, the rules of moral grammar constrain the causal structures of moral theories while still allowing a wide variety of different moral theories to exist.

The unequal vividness explanation accounts for the widespread moral agreement that Mikhail and Hauser describe. It does so while explaining the nature of the universal moral grammar so that it doesn't have to be taken as a primitive and unexplained feature of our psychology. If it's simply part of human psychology that we more vividly imagine the intended than the merely foreseen and have more violent passions towards more vividly represented things, this would explain the universality of intuitions supporting Double Effect. It would be nice if Chomsky's syntactic rules could be explained in terms of further psychological states or processes showing why syntax has to be this way.[36] Then we wouldn't have to take them as primitive psychological facts, unexplained by anything else. I explain some features of the universal moral grammar in terms more general than those of morality. Some elements of the unequal vividness explanation – imagining the intended more vividly than the merely foreseen and having stronger emotions towards what we represent more vividly – are observed in non-moral phenomena just as much as moral phenomena. Others –

[34] Marc D. Hauser, Liane Young and Fiery Cushman, 'Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions', *Moral Psychology*, vol. 2: *The Cognitive Science of Morality: Intuition and Diversity*, ed. Walter Sinnott-Armstrong (Cambridge, Mass., 2008), pp. 107–45.

[35] John Mikhail and C. Sorrentino, 'Toward a Fragment of Moral Grammar: Knowledge of the Principle of Double Effect in Children Ages 8–12', poster presented to the Society of Research in Child Development (1999); John Mikhail, C. Sorrentino, and Elizabeth Spelke, 'Toward a Universal Moral Grammar', *Proceedings: Twelfth Annual Conference of the Cognitive Science Society*, ed. M. A. Gernsbacher and S. J. Derry (Mahwah, NJ, 1998), p. 1250.

[36] Chomsky himself has tried to develop such explanations. See Noam Chomsky, *The Minimalist Program* (Cambridge, Mass., 1995).

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

strong emotions being able to shape our moral judgement and taking the first-personal perspective when judging others – are general facts about moral judgement that we have independent reason to accept. In offering a simpler psychological explanation where others might posit primitive and unexplained psychological facts, the unequal vividness explanation marks a theoretical advance.

BENNETT'S PHENOMENON AND OTHER TROLLEYOLOGY

When presenting the doctrine of Double Effect, I didn't define the notions of intended and foreseen consequences. Trying to do so reveals an interesting phenomenon noted by Jonathan Bennett: on an intuitive picture of the difference between intended and foreseen consequences, all the harms in canonical Double Effect cases come out as foreseen. I'll lay out this intuitive picture of the intended/foreseen distinction and then illustrate Bennett's phenomenon. The unequal vividness explanation helps us deal with Bennett's phenomenon, equipping us to maintain this intuitive view of intending while explaining why we see Double Effect's distinctions between harms as we do. Once we see this, we can also see how it suggests a way of explaining results from a variety of other trolley cases.

Here's the intuitive position that I'll call the 'straightforward view' of the distinction between intended and foreseen consequences. For a consequence to be intended, it must be in the series of steps beginning with our action and concluding with the end. Events outside the series leading from the action to the end aren't intended. If we believe that they'll result from the action, they're merely foreseen but not intended. I call this the 'straightforward view' not just because it's a simple and intuitive view, but because it treats what's intended as a sequence going straight forward from the action to the desired state, without branching off in directions that don't raise the probability of the goal. Events on the branches don't help agents achieve their goals, so they aren't intended.[37]

Even in Double Effect cases usually described as ones where the agent intends to harm someone as a means, the straightforward view presents the intention as stopping just short of the actual harm, treating the harm as a merely foreseen consequence. Bennett's phenomenon is that these harms are intuitively regarded differently from other foreseen harms, even though they are foreseen harms according to the straightforward view. On the straightforward view, a utilitarian who throws the fat man intends that the man's body

[37] This notion of branching off is nicely illustrated by the act-trees in John Mikhail, 'Rawls' Linguistic Analogy' (PhD dissertation, Cornell University, 2000).

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

will block the trolley, since this has the effects that are important to achieving her goal. But the fat man's death, independently of his blocking the trolley, has no such consequences, so she doesn't intend his death.

It's natural to regard the fat man's death as extraneous to the utilitarian's intention, as the straightforward view suggests. Consider what she might say if it turned out, much to her surprise, that the fat man was a superhero with the power to turn himself into titanium and thus block the trolley's progress without being harmed. If you ask the utilitarian 'Was that contrary to what you intended?' after she sees the fat man turn himself into titanium and block the trolley, she'll rightly say 'No'. While she didn't expect such a wonderful thing to happen, and certainly didn't intend it, it wasn't contrary to her intention. The fat man turning himself into titanium leaves her series of intended means entirely intact. Now consider a scenario where the fat man has a different superpower. He turns into a mist, allowing the trolley to pass through and kill the five men behind him. Now suppose you ask the utilitarian 'Was that contrary to what you intended?' She'll rightly say 'Yes'. She intended the man's body to block the train, saving the five men behind him. Her goal can't be achieved if the man turns into a mist. The series of intended means leading from her action to its goal is disrupted.

Understanding what was intended in accordance with the straightforward view would, as McIntyre writes, 'drastically limit the prohibitive force of [Double Effect]'.[38] The impermissibility of terror bombing and throwing men at trolleys arising from Double Effect is supposed to extend specifically to agents who intend things that the straightforward view treats as stopping just short of creating harms. The intuitions supporting Double Effect present the harms in these cases differently from the harms of strategic bombing and switching the trolley so it hits only one person. This raises a question: why don't bad consequences like the fat man's death that the straightforward view of the intended/foreseen distinction classifies as merely foreseen seem like other merely foreseen consequences?

The unequal vividness explanation answers this question. When we imagine what we'd intend if we made the utilitarian choice in the trolley case, it's hard not to imagine the fat man's pain and death with the same full vividness as we imagine what the straightforward view says we intended – namely, that his body would block the train. The nearly-but-not-quite-intended consequences of his pain and bloody death are part of the same scene as the fully intended consequences of his body

[38] McIntyre, 'Doing Away With Double Effect', p. 234.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

blocking the train. In fact, it's the most horrific part of that scene, so it draws our attention. The merely foreseen bad consequences in the case where we throw the switch, diverting the trolley so that it kills only one, are more distant from the things we're forced to imagine as we initially form or later contemplate our intention, so we can avert our imaginative gaze from them. But if we intend to block the train with the body of a human being who will thus die a gruesome death, there's no room for averting our imaginative gaze from the death as we imagine the blocking.

The straightforward view doesn't itself distinguish intended and foreseen harms in a way that accords with our Double Effect intuitions. But when combined with the unequal vividness explanation, it shows us where the intuitive line is. The harms we intuitively regard as worse include those that aren't intended, but are so close to the things we intend that we can't help but imagine them vividly when we imagine what we intend. To switch examples, even if the terror bomber doesn't intend to kill civilians, but merely to blow them apart, he'll naturally imagine killing them when he imagines blowing them apart. The strategic bomber, meanwhile, can easily imagine the destruction of military targets while less vividly imagining civilian deaths. When we put ourselves in the agents' shoes and consider doing what they did, this difference leaps out at us. And that's why we feel differently about the cases distinguished by the Doctrine of Double Effect.

Here one might wonder about how to explain other interesting trolley cases. Thirty-seven per cent of subjects think it's permissible to throw a switch dropping a fat man onto the tracks to save the five men (compared to 10 per cent willing to push the fat man). Sixty-two per cent think it's permissible to throw a switch that will collapse a footbridge onto the track so that the footbridge can block the trolley and save the five men, though someone standing on the footbridge will die in the fall.[39] How should we explain these results? I suggest that the vividness with which we imagine the one person's death, whether produced by its being intended or by other factors, produces the judgements of impermissibility in these cases. If we more vividly imagine things near us that we touch as well as see, that explains why we're less willing to push a fat man with our own hands than to throw switches that drop him. And if we don't have to imagine a man dying in blocking the trolley, but a footbridge blocking the trolley instead (though a man dies as the footbridge falls), the death recedes somewhat from the chain of intended events that we have to imagine directly, but not quite as much as in the case where we switch the trolley to a side track. Schaich

[39] John Mikhail, 'Universal Moral Grammar: Theory, Evidence, and the Future', *Trends in Cognitive Sciences* 11 (2007), pp. 143–52, at 149.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

Borg's finding that more colourful descriptions of the same harms raised impermissibility judgements in Double Effect cases supports the contention that a great deal of our thinking about the permissibility of bringing about various harms is the result of systematic differences in how vividly we represent them.

THE UNRELIABILITY OF UNEQUAL VIVIDNESS

Having explained the appeal of Double Effect in terms of the unequal vividness with which we imagine intended and foreseen harms, I'll now argue that this explanation reveals why we shouldn't ascribe moral significance to the distinction. As I'll argue, when the effects of unequally vivid representations upon our desires determine our decisions about what to do, we are usually mistaken.

Some philosophers may consider it obvious that the greater appeal of one option over another caused by unequal vividness gives us no reason to choose it. They might argue as follows. The appeal that unequal vividness gives to an option doesn't correspond to anything of real moral significance. So when the distinction between harms intended and harms merely foreseen plays a decisive role in our decision-making, we decide on the basis of something that fails to track the merits of our options. When we see that some principle fails to track the merits of our options, we see that it shouldn't be relied on for normative guidance. So we shouldn't rely on Double Effect.[40]

This argument isn't enough. The claim that unequally vivid representation doesn't correspond to anything of moral significance is unsupported, and the argument thus fails to address the possibility that unequal vividness informs us about an important normative feature of the world. Perhaps the lesser vividness of merely foreseen outcomes is a veridical perception of their lesser genuine moral significance to us as agents, just as the faintness of faraway sounds tells us something true about their distance from us. Their lesser vividness and emotional impact might correctly represent the lesser wrongness of our bringing them about, even though they may be equally bad when considered from a different point of view from ours. To overcome this defence of Double Effect, we should consider other cases in which unequally vivid representations affect our desires, and see if these unequally vivid representations can plausibly be regarded as making us aware of some important normative feature of the world.

[40] I take this argument to be subject to the most severe problem that Berker presents for the best neuroscientific argument against deontology he considers: that instead of drawing on the force of any empirical results, it simply rests on armchair identification of some factors as morally relevant and others as irrelevant (Berker, 'Normative Insignificance', pp. 325–6).

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

When unequal vividness is responsible for our decisions, so that we would have acted differently if we represented the options with equal vividness, we typically err.[41] Decisions where we sacrifice greater long-term benefits for lesser goods that are immediately available often fit this pattern. The recovering alcoholic acts irrationally when he drinks from the bottle of whisky he sees in front of him at the party, which presents the pleasures of drinking to him more vividly than he is presented with the more distant costs of relapsing into alcoholism. The football player acts irrationally when he hits the opponent who taunts him, drawing a penalty that creates obstacles for his team and turns his fans' admiration into anger. I act irrationally when I stay up too late watching silly videos on YouTube, which are more vividly presented to me at 3 a.m. than are the benefits of waking up with a good night's sleep tomorrow and getting work done. These are all cases in which, very broadly speaking, one pursues a lesser nearby good at the expense of a greater distant good. One pursues the short-term satisfactions of drinking, striking a noxious opponent or wasting time on minor amusements at what one's own reflective judgement would portray as greater long-term costs.

Each of these cases can be explained by how the more vivid option produces more violent passions than the less vivid one. The agents might have acted differently if the vividness of the options before them were equalized. If the recovering alcoholic had been presented with a vivid image of others' contempt for him after his relapse into addiction, he might have regarded the idea of drinking with horror. If the football player had been presented with a vivid image of the opposing team celebrating victory after his penalty as his team mates walked dejectedly off the field, he would have kept his cool. If I had been presented with a vivid image of the embarrassment of being caught in Q&A by a devastating objection, I would have closed Firefox and gone to bed so that I could get more work done the next day. But without such images, their greater good excites only calm passions, and their violent passions drive them irrationally to prefer the lesser but more vividly represented good. While these ways of modifying the cases involve counterfactually strengthening the vividness with which agents imagine or sense the distant options so that they represent them as vividly as the nearby options, we can get the same results by

[41] Many moral theorists have noted the importance of equally vivid representation of options for good decision-making. See R. M. Hare, *Moral Thinking* (Oxford, 1981); Roderick Firth, 'Ethical Absolutism and the Ideal Observer', *Philosophy and Phenomenological Research* 12 (1952), pp. 317–45. For a discussion of how unequally vivid representation of options leads to systematically irrational action, see Neil Sinhababu, 'The Humean Theory of Practical Irrationality', *Journal of Ethics & Social Philosophy* 6 (2011), pp. 1–13.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

weakening the vividness of the nearby options to make them look like the distant options. This is what happens when people consider their options in a cool moment, far away from the vivid images that will rouse their passions. The day before the party, the recovering alcoholic judges that the right choice would be not to drink. The night before the game, the player judges that the right choice would be to ignore the taunts and focus on winning. In the afternoon, I plan to get to bed on time, not procrastinate on the internet. Equal low vividness can give rise to good decisions, just as equal high vividness can.

The point of these cases is that when unequal vividness affects our desires and thus is decisive in driving our decisions, we typically do the wrong thing. If it's true, as I've argued in the previous section, that the appeal of Double Effect is merely an artefact of unequal vividness, actions based on this principle are similarly likely to be wrong. Our sense that it's wrong to throw the fat man in front of the trolley or engage in terror bombing rather than strategic bombing involves the same sort of error that is present when the alcoholic pours himself a drink, the football player hits his taunting opponent, or the procrastinator wastes valuable time on trivial amusements. Tempted by nearby and vividly represented goods, a wise decision-maker may make an effort to imagine the long-term costs. Perhaps this would also be wise in ethics, and we would be better moral decision-makers if we made an effort to imagine the lost lives of all five men we could save from the runaway trolley as vividly as we imagine that of the one we might kill.

The neurological processes implicated in Double Effect overlap with those implicated in preferring lesser near-term goods to greater long-term goods, suggesting that some similar psychological process operates in both cases. The PC, PCC and orbitofrontal cortex, which are activated in cases like those of Double Effect where deontological intuitions are common, also seem to be part of a neural system responsible for the pursuit of lesser immediate gains over greater distant gains.[42] If these regions are implicated in the more vivid representation of intended harms as previously suggested, this provides neuroscientific support for a unified unequal vividness explanation of preferring one's lesser near-term good and the intuitiveness of Double Effect.

Here I've been using bad decisions resulting from unequally vivid representation to argue that moral judgements produced by unequally vivid representation are mistaken. One might object that decisions to

[42] S. M. McClure et al., 'Separate Neural Systems Value Immediate and Delayed Monetary Rewards', *Science* 306 (2004), pp. 503–7; S. M. McClure et al., 'Time Discounting For Primary Rewards', *The Journal of Neuroscience* 27 (2007), pp. 5796–804; C. S. Sripada et al., "The Neural Correlates of Intertemporal Decision-Making', *Human Brain Mapping* 32 (2011), pp. 1637–48.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

do something and judgements about what we should do are different things, and ask why systematically bad decisions under some conditions should lead us to suspect that judgements under those conditions are systematically false. My answer is that even if the irrational agents didn't get around to explicitly judging that they should do as they did, they had misleading sensory or imaginative representations that disposed them to make such judgements. While decision and judgement can come apart in cases of akrasia, these cases are atypical. As Richard Holton has argued, it's more common to have 'judgement shift' where giving in to temptation is accompanied by a judgement that it's okay to do so this time.[43] If something systematically leads to wrong action, it's likely to lead systematically to bad judgements about which action is right. If unequally vivid representations can make us weigh our moral reasons poorly in Double Effect cases, leading to the wrong decision as they do in so many other cases, it would be surprising if they couldn't lead us to false judgements about which decision one should make. At least in usual cases, moral decision and moral judgement are tied together so tightly that a force creating error in one will create error in the other. Even someone who tries to block the attack on Double Effect by denying the link between decision and judgement will have to explain why actions in accordance with Double Effect are right, when they result from processes that systematically generate bad decisions. Defenders of Double Effect can't say that judgements according with their principle are correct about the moral facts, while decisions according with it are irrational. The judgements, after all, are about the decisions, and the reasons involved are moral reasons.

Fortunately, the effects of unequal vividness often aren't decisive. So, not all decisions involving some degree of unequally vivid representation of the options are mistaken. Our stronger desire for one good may just overpower our weaker desire for a second good, even if the latter is more vividly represented. And in some cases, we represent our more-desired good more vividly, adding extra force to what is the rational decision anyway. It's only when equal vividness would have resulted in one decision but unequally vivid representations tip the balance and make us do otherwise that we systematically act irrationally. If the unequal vividness explanation is correct, decisions and judgements in accordance with Double Effect will fall into this systematically erroneous class.

Surely unequal vividness sometimes saves us from error by counterbalancing other forces leading to poor decisions, such as overconfidence about attaining uncertain future goods. This might

[43] Richard Holton, *Willing, Wanting, Waiting* (Oxford, 2009), ch. 5.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

make acting on unequal vividness a reliable heuristic when these other unreliable processes are at work. Defenders of Double Effect as a genuine moral principle can't rest easy with this defence of unequal vividness, though. They want to apply the doctrine even to cases when we know what's going to happen, where the features of ordinary situations which might make unequal vividness reliable are missing. If unequal vividness leads to good decisions merely as a counterweight to overconfidence about future goods, that doesn't support relying on it when we know what will happen. We wouldn't regard overconfidence about future goods as a reliable source of good decisions just because it sometimes saves us from error by counterbalancing unequal vividness. If unequal vividness and overconfidence about future goods only lead us right as counterweights to each other, that leaves them in an equally dubious position.

As I've argued, the same mechanisms of unequally vivid representation and violent passion that make Double Effect seem plausible also make irrational neglect of long-term goods seem choiceworthy. With both Double Effect and irrational neglect of long-term goods, one option is represented more vividly than another, making our passions for it more violent, and leading us to choose it when we otherwise would not have. There are minor differences. With Double Effect, the more vivid representations arise from a structural feature of deliberation, making us imagine things intended more vividly than things merely foreseen. With neglecting our long-term good, the more vivid representations arise from the more immediate availability of our short-term good or its sensed physical presence. But differences in the causes of unequal vividness are compatible with the similarity that is essential to my argument – that unequal vividness, whatever its causes, leads our decisions and judgements astray.

Representing our options in our senses or imagination as we deliberate is an important part of how we determine what we have most reason to do. This holds both when we're explicitly considering moral reasons and when we don't see our reasons as morally charged. The backsliding alcoholic, the enraged football player, and the procrastinator demonstrate how unequal vividness leads to mistakes. If following Double Effect is similarly produced by unequal vividness, it's likely to be another way to make mistakes.

THREE RESPONSES

Now I'll consider three ways for defenders of Double Effect to respond to my argument. The argument has two premises – a psychological story on which unequal vividness moves us to accept Double Effect and

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

an unreliability claim according to which decisions in which unequal vividness is decisive are likely to be mistaken.

The most impressive response would be to deny the first premise, developing a better-supported psychological story on which the appeal of Double Effect was produced by a more reliable process than unequal vividness. I've argued that unequal vividness explains a wide range of phenomena, including the neuroscientific data described by Greene, Klein, Schaich Borg, and Kahane, Hauser and Mikhail's data about the universality of Double Effect intuitions, and Bennett's phenomenon. If defenders of Double Effect produce a better psychological story on which their doctrine is the result of reliable processes, they'll not only defeat my argument, but move psychology forward. I won't anticipate and discuss particular ways in which they could develop such a psychological picture, because I don't have a good sense for which way their theoretical creativity might take them. I hope the unequal vividness explanation is good enough to make this a non-trivial task.

Another response would be to deny the second premise and defend the general reliability of unequal vividness, both where it leads us to neglect greater distant goods and in cases of Double Effect. One might endorse reckless neglect of distant goods and reap the benefit of defending Double Effect by claiming that unequal vividness is a reliable way of discovering normative facts. But I think the irrationality of neglecting greater goods because they're distant is too obvious to make this attractive. If Double Effect and neglecting greater distant goods in favour of lesser nearby goods are in the same boat, this will sink Double Effect rather than saving wanton neglect of the future.

Another response would be to grant that Double Effect and neglect of greater distant goods issue from unequal vividness, and accept that the latter is irrational, but insist that this doesn't justify rejecting Double Effect. This response would involve accepting that unequal vividness often misleads us, but rejecting any version of the unreliability claim strong enough to license the conclusion. As many defenders of Double Effect may be inclined towards this option, I'll criticize it at greater length.

I concede that there are cases in which a move like this could work. Consider the generally unreliable process of wishful thinking. There may be domains in which wishful thinking systematically leads to true beliefs – perhaps in determining whether things will go well for the protagonists of children's books. Perhaps one shouldn't give up one's belief that Curious George will get home safely, or that Brother and Sister Bear will be reconciled, just because wishful thinking produced it. Beliefs so caused might be justified by the fact that the children's books are written to accord with childlike wishes.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

But in general, when one believes something because of a generally unreliable process like wishful thinking, and one has no further evidential support, one should give up the belief. On most topics, like current events, investing, science and romance, if someone claims that another person's belief is wishful thinking, we take this as a criticism requiring a rebuttal. To defend the believer, we might cite genuine evidence for the claim, suggesting that the belief is in fact supported by something epistemically better than wishful thinking. Sometimes, as with children's books, we might argue that wishful thinking is actually reliable because of specific features of the case. We offer these defences because we know that without them the claim that a belief was produced merely by the wildly unreliable process of wishful thinking is an undercutting defeater for it. Unless we can demonstrate that the defeater doesn't apply or is itself defeated somehow, there is no justification for believing.

The unreliability of unequal vividness, in cases like those of neglecting our greater long-term good, leaves it as poor a guide to the weight of our reasons as wishful thinking is about the future. Defenders of Double Effect thus can't just accept the unequal vividness explanation and accept that unequal vividness is generally unreliable. They need to deny the unequal vividness explanation, endorse the reliability of unequal vividness, or explain why unequal vividness causes us to weigh our reasons rightly in the case of Double Effect while it fails elsewhere.

A METHODOLOGICAL CONCLUSION

In 'The Normative Insignificance of Neuroscience', Berker lays out several arguments offered by Singer and Greene which try to undermine deontological intuitions by appealing to neuroscientific results, including the results that I've cited. I agree with many of his criticisms of these arguments. At the end of his paper, he suggests a way in which arguments from neuroscience could be useful. I'll conclude by examining the strategy he suggests, explaining how my argument has the feature that makes it successful. Much of my argument has relied on folk-psychological investigations into the processes underlying our intuitions rather than neuroscientific ones. But at times I've drawn some support from neuroscience, and at any rate arguments against the normative significance of neuroscience could be generalized into arguments against the normative significance of other empirical methods. I'll explain why the arguments I've offered here are better than those Berker rejects.

Berker presents a 'best-case scenario' of how neuroscience could 'play a more direct role in our theorizing about the evidential status of moral

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

intuitions'.[44] In this scenario, 'We notice that a portion of the brain which lights up whenever we make a certain sort of obvious, egregious error in mathematical or logical reasoning also lights up whenever we have a certain moral intuition.'[45] Presumably the error needs to be mathematical or logical because our knowledge of mathematics and logic is very secure. If deontological intuitions and errors in arithmetic arose through the same process and consequentialists argued that this revealed the unreliability of deontological intuitions, deontologists couldn't plausibly respond that the process was in fact reliable and that we should revise arithmetic. The principle explaining why this scenario counts as a best-case scenario for neuroscientific arguments debunking deontology is something like the following: if a class of intuitions arises from a process that we know on independent grounds to be systematically unreliable in producing correct judgements, we shouldn't rely on those intuitions in our theorizing.

While pursuing one's lesser near-term good at the expense of one's greater long-term good may not be as obviously wrong as making errors in arithmetic, we still have good independent grounds for regarding it as a mistake. The irrationality of backsliding alcoholics, enraged football players and procrastinators is obvious enough to diminish the reliability of whatever psychological process is driving the decisions here and in similar cases. The epistemic profile of my argument is roughly the same as in Berker's best-case scenario. In both cases, we identify the right and wrong answers in some domains (mathematics or practical rationality) from the armchair. Then we use empirical methods to determine how people come to the wrong answers in those domains, and also how people arrive at some moral principle. We discover that intuitions favouring the moral principle arise through the same process that produces wrong answers. Seeing that our intuitions favouring the principle arise through a highly unreliable process, we discount these intuitions in our theorizing.

Berker continues:

> If . . . we come to see that the moral intuition in question rests on the same sort of confusion present in the mistaken bit of mathematical/logical reasoning, then of course we should discount the moral intuition, but in that case the neuroscience isn't playing a direct justificatory role. Again, we might not have thought to link the moral intuition to that sort of mathematical/logical blunder if we hadn't known the neuroscientific results; but again, once we do link them, it seems that we do so from the comfort of an armchair, not from the confines of an experimental laboratory.[46]

[44] Berker, 'Normative Insignificance', p. 328.
[45] Berker, 'Normative Insignificance', p. 329.
[46] Berker, 'Normative Insignificance', p. 329.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press

This understates the utility of neuroscience. The processes driving the formation of various moral intuitions won't always be obvious. Whether some moral intuition results from one process which is agreed to be reliable or another which is agreed to be unreliable may be a point of contention between moral theorists who accommodate the intuition and those who don't. In these cases, hard evidence about the nature of the processes producing the intuition will help us settle the psychological debate. Simple folk-psychological investigations of our own reasoning have much to offer, and I've relied heavily on them. But as neuroscience progresses, it may become our best way of determining which process produced a particular judgement. By helping us separate our processes of belief-formation into classes that can be properly assessed for reliability, neuroscience and all our other modes of investigating how we think can help us make progress in ethics.[47]

neiladri@gmail.com

[47] This article was improved by comments at Waikato University, the University of New South Wales, Macquarie University, the University of Western Australia, Michigan State University, the University of Illinois at Chicago and the Ian Ramsay Centre at Oxford. Travel to present this article as a paper was enabled by generous support from the National University of Singapore. Andrew Higgins, Colin Klein, Selim Berker, and an anonymous referee for *Utilitas* helped me with a variety of issues and pointed me to important empirical literature. Encouragement from Jesse Prinz and Walter Sinnott-Armstrong was also greatly appreciated.

https://doi.org/10.1017/S0953820812000362 Published online by Cambridge University Press