*Manuscript accepted for publication and forthcoming as part of an Open Peer Commentary on Joshua May’s* Regard for Reason in the Moral Mind *in* Behavioral and Brain Sciences*, subject to peer review and/or editorial input by Cambridge University Press and/or the journal’s proprietor. Please cite final version, which will be available at* [*https://www.cambridge.org/core/journals/behavioral-and-brain-sciences#*](https://www.cambridge.org/core/journals/behavioral-and-brain-sciences)*. Copyright owned by Cambridge University Press*.

**Analyzing debunking arguments in moral psychology:**

**Beyond the counterfactual analysis of influence by irrelevant factors**

Joanna Demaree-Cotton

Department of Philosophy, Yale University

joanna.demaree-cotton@yale.edu

**Abstract:** May assumes that if moral beliefs are counterfactually dependent on irrelevant factors, then those moral beliefs are based on defective belief-forming processes. This assumption is false. Whether influence by irrelevant factors is debunking depends on the mechanisms through which this influence occurs. This raises the empirical bar for debunkers and helps May avoid an objection to his Debunker’s Dilemma.

In Chapter 4 of *Regard for Reason in the Moral Mind (RRMM),* May tackles sweeping debunking arguments that aim to show that ordinary moral beliefs are not justified because they are inappropriately influenced by morally irrelevant factors, such as incidental disgust, how a moral scenario is worded or “framed,” or whether or not an agent uses “personal force” in order to bring about a harm. May argues that, for the debunker’s argument to succeed, they need to identify an influence on belief that is both substantial (empirical premise) and defective (normative premise). A defective influence on belief is an influence on belief that, if substantial, renders that belief unjustified (such as wishful thinking). However, May argues that debunkers are faced with the following dilemma: Either the irrelevant factor in question is not a “substantial” influence on moral belief, thus undercutting the empirical premise, or the debunker has identified a substantial influence on belief that is not defective, thus undercutting the normative premise.

Throughout his book, May slides between two ways of presenting the would-be debunker’s empirical premise, sometimes writing in terms of:

1. Whether or not moral beliefs are substantially influenced by epistemically defective *factors* (that is, morally irrelevant factors)

And sometimes writing in terms of:

1. Whether or not moral beliefs are substantially influenced by epistemically defective *processes*

It is fairly clear what it is for a moral belief to be the product of an epistemically defective process. Wishful thinking, motivated reasoning, and paranoid inferences are all examples of belief-forming processes that result in unjustified beliefs (e.g. *RRMM,* p. 85).

But what is it, exactly, for a moral belief to be substantially influenced by an irrelevant factor? May seems to rely on an analysis in terms of *counterfactual dependency.* On this view, moral beliefs are substantially influenced by an irrelevant factorif the agent would have formed a different moralbelief – a belief with a different polarity or valence – had that factor been absent (*RRMM*, p. 213). For example, my moral belief that an agent is doing something morally wrong is substantially influenced by incidental disgust in this sense if, *had* I not experienced incidental disgust, I *would not* have believed that the agent is doing something morally wrong. Similarly for framing: If one accepts this counterfactual analysis of the debunker’s empirical premise, then “what a debunking argument requires” […] is “that people regularly tend to lose their belief or change its content” if the framing of a moral problem is altered (*RRMM*, p.90).

It seems that the reason that May is happy to slide between talk of epistemically defective processes and talk of dependence on irrelevant factors is that he assumes that if a moral belief is counterfactually dependent on a morally irrelevant factor, then it is the product of a defective process – one that, like wishful thinking, guesswork, or motivated reasoning, is unreliable, insensitive to evidence, or otherwise yields unjustified beliefs (*RRMM*, p.85). This assumption is tempting and is widely held by debunkers (e.g., Sinnott-Armstrong 2008) and by anti-skeptical defenders of moral judgment alike. Indeed, I have explicitly outlined and relied on such a counterfactual analysis of the threat of irrelevant factors elsewhere (Demaree-Cotton 2016).

The trouble is that this assumption – the assumption that if a moral belief is counterfactually dependent on an irrelevant factor, then it is the product of a defective process – is false. Moreover, it concedes too much to the debunker.

The assumption is false because counterfactual dependence per se tells us exceedingly little about what the actual psychological process was that led the agent to form the moral belief that they did, let alone whether or not that process is a defective one. There are a number of reasons for this. One is that an irrelevant factor can influence moral belief precisely because it influences what kind of belief-forming process an agent engages in. Imagine an experimenter induces incidental feelings of anger, and then presents me with a moral vignette and asks whether the main character is doing something morally wrong. Perhaps because of the incidental anger, I feel engaged by the task – the topic of morality feels, right now, like something interesting and important, worthy of deep reflection – and consequently I carefully reflect on details of the scenario, weighing up whether or not there is a suffering victim, whether or not this suffering constitutes an injustice, and contemplating the main character’s role in the production of that injustice. On the basis of these considerations I conclude the main character did something morally wrong. This is, I take it, a canonical example of a justification-conferring psychological process. The resulting belief is justified. Moreover, it would remain justified even if, had some irrelevant factor been different – for example, had I not just gone through the anger induction – I would have engaged in a different type of belief-forming process and formed a different belief as a result. This fact cannot in itself bear on whether or not my actual belief is justified, because it does not bear on what the actual psychological process was that led to that belief.

This is so even if, counterfactually, I would have engaged in a defective belief-forming process. Imagine that instead of anger the experimenter used a mood manipulation to enhance feelings of cheerfulness. Moreover, if she had, I would have taken little interest in the moral dilemma presented to me; while daydreaming about my weekend plans, I would have absent-mindedly formed the belief that what the main character is doing is probably perfectly morally permissible. This is not a reliable, rational way of forming moral beliefs, and the resulting belief would not be justified. Still, this counterfactual fact about me in no way impugns the careful reasoning I actually engaged in after the anger induction and the justificatory status of the belief I formed on the basis of that reasoning.

Another possibility is that a morally irrelevant factor affects, not what *type* of belief-forming process you engage in (e.g. reasoning based on evidence versus motivated guesswork), but merely what subset of all of the relevant evidence you (1) notice, and (2) pay attention to and weigh when arriving at your belief (see Avnur & Scott-Kakures 2015, on “positional” influences of irrelevant factors on belief). To return to the example above, perhaps my feelings of incidental anger remind me of cases of injustice I experienced in the past, and consequently I am especially sensitive to aspects of the moral situation described in the vignette suggesting that the victim is suffering an unjust harm – aspects of the moral situation that influence my judgment and that I might not have noticed had I been in a different mood.

Of course, this is a purely hypothetical example, and not an empirically-grounded one. This may well not be the mechanism by which incidental anger can be expected to influence moral belief. But this is exactly the point. For the debunker’s argument to work, it is *not sufficient* for them to cite studies showing morally irrelevant manipulations have a significant effect on moral beliefs, even if the effect sizes in question are large. They must bring to bear theoretical interpretations of those effects on moral beliefs that give us reason to think that the proximate psychological mechanisms driving those moral beliefs are indicative of defective, non-justification-conferring psychological processes.

This also shows why May’s slide between talk of a moral belief being “substantially influenced” by irrelevant factors and talk of a moral belief being “mainly based on” irrelevant factors is liable to mislead in a way that does his argument a disservice. We normally use the term “basis” to pick out, not just *any* substantial causal influence on belief, but a cause that is a crucial part of the psychological process leading to the belief (typically, a piece of the agent’s evidence). For example, in the hypothetical anger induction example, my moral belief was counterfactually dependent on whether or not I had undergone an incidental mood induction; but my moral belief was *based* (in a psychological, evidential sense) observations I made about injustice. To conflate causal influence with psychological basis risks seeing defective belief-forming processes where there are none.

The final difficulty I want to outline regarding the counterfactual analysis is not just that it is false, nor just that it obscures the need for appeals to details of psychological process and mechanism, but that it concedes too much to the debunker – potentially in ways that threaten to weaken May’s anti-skeptical arguments. A really important part of May’s anti-skeptical argument is his motivation of the Debunker’s Dilemma and the idea that *any* empirically-based debunking argument will most likely face a tension between establishing that moral beliefs are substantiallyaffected by some source and establishing that that source renders the belief unjustified. Why think the dilemma will generalize? According to May, since there are *many different* influences on belief – some of which are defective, and some of which are non-defective – it is unlikely that any *particular* defective influence will have a substantial effect on a large class of our moral beliefs (*RRMM*, pp.103–104).

Furthermore, he argues that it is unlikely that lots of individually insubstantial defective influences (such as morally irrelevant factors) will add up to have a substantial effect on a large class of moral beliefs in a way that is debunking, because individually insubstantial *appropriate* influences on beliefs (such as morally relevant factors) can add up in exactly the same way (*RRMM*, p.229).

But it is unclear why this would be true if we thought that mere counterfactual dependence on irrelevant factors counted as a defective influence on belief, where that influence is specified in a way that abstracts away from psychological process. Any moral belief we form has been affected by countless irrelevant factors in this purely counterfactual sense, insofar as the development our evidence (including our background beliefs and our moral convictions) and what exact belief-forming process we engage in (including what evidence we consider and what type of thinking or reasoning we use to form the belief) is counterfactually dependent on an innumerable number of irrelevant factors, from our evolutionary history, to where we were born, to our socioeconomic status, to our health, to our mood, to whether or not we happened to have faced a similar moral problem previously, and so on. The number of factors that are morally relevant to the problem at hand that could influence us, by contrast, is necessarily restricted.

This problem goes away if we recognise that only a specific way of being influenced by irrelevant factors is pertinent to assessing the justificatory status of moral belief – namely, being influenced in such a way that leads you to engage in a particular kind of defective belief-forming process (e.g., because the factor in question tends to produce motivated reasoning, or because we tend to form moral beliefs that use that irrelevant factor as a heuristic, although the heuristic in question is unreliable). If we focus psychological dependence on irrelevant factors, rather than a more general sense of counterfactual dependence, then it is much more plausible that small appropriate influences are going to stack up against small inappropriate influences on belief.

In summary, I have presented an argument that is both critical of an assumption that May makes in *Regard for Reason in the Moral Mind,* but one that he should welcome if he wishes to defend ordinary moral belief. Counterfactual dependence on irrelevant factors does not matter. What this shows is that, in order to succeed, would-be debunkers have to meet a much more stringent empirical premise than May has allowed. May is absolutely right that statistical details matter (*RRMM*, p. 229). It matters how big of an effect irrelevant factors have on our moral beliefs. But it also matters *what kind* of an effect irrelevant factors have on our moral beliefs – the details of the psychological processes through which irrelevant factors come to affect our moral beliefs are of crucial importance when assessing the merits of debunking arguments in moral psychology.

References

Avnur, Y., & Scott-Kakures, D. 2015. “How Irrelevant Influences Bias Belief.” *Philosophical Perspectives,* 29: 7-39.

Demaree-Cotton, J. 2016. “Do Framing Effects Make Moral Intuitions Unreliable? *Philosophical Psychology* 29(1): 1-22.

May, J. 2018. *Regard for Reason in the Moral Mind.* New York, NY: Oxford University Press.

Sinnott-Armstrong, W. 2008. “Framing Moral Intuitions.” In *Moral Psychology, Vol. 2, The Cognitive Sicence of Morality: Intuition and Diversity,* ed. W. Sinnott-Armstrong (pp.47-76). Cambridge, MA: MIT Press.