

Moral Nihilism—So What?*

Lewis Williams

Edward Elliott and Jessica Isserow argue that it is not usually in the best interests of ordinary human beings to learn the truth of moral nihilism. According to Elliott and Isserow, ordinary human beings would suffer costs from learning the truth of moral nihilism that are unlikely to be fully compensated for by any benefits. Here I provide reasons to doubt that ordinary human beings would suffer costs from learning the truth of moral nihilism and present a dilemma for Elliott and Isserow's view.

I. INTRODUCTION

In “Don’t Stop Believing (Hold onto That Warm Fuzzy Feeling),” Edward Elliott and Jessica Isserow argue that if an ordinary human being had the option to learn the truth of moral nihilism, they rationally ought to reject the offer.¹ In doing so, they argue that learning the truth of moral nihilism would preclude the ordinary human being from experiencing “warm fuzzy feelings” (*wff*'s) for acting rightly, and that no benefits that arise from learning the truth of moral nihilism can be expected to offset these costs.

In this article, I provide reasons to doubt that ordinary human beings would suffer costs from learning the truth of moral nihilism, and I present a dilemma that this poses for Elliott and Isserow's view. I begin in Section II by expounding the structure and commitments of Elliott and Isserow's argument. In Section III, I argue that ordinary human beings can still be expected to experience the *wff*'s identified by Elliott and Isserow, even after coming to learn the truth of moral nihilism. In Section IV, I explore the

* For helpful comments, thanks to Guy Kahane, Kyle van Oosterum, Edward Elliott, Jessica Isserow, and an associate editor of *Ethics*. This work has been supported by the Faculty of Philosophy and Linacre College at the University of Oxford via the Ryle / T. H. Green Moral Philosophy Scholarship and by the Forethought Foundation via a Global Priorities Fellowship.

1. Edward Elliott and Jessica Isserow, “Don’t Stop Believing (Hold onto That Warm Fuzzy Feeling),” *Ethics* 132 (2021): 4–37, henceforth cited as “DSB.”

ramifications for Elliott and Isserow's wider argument and present a dilemma for their view. I conclude that the evidence provided by Elliott and Isserow undersupports their thesis that ordinary human beings rationally ought not to inquire after the truth of moral nihilism.

II. ELLIOTT AND ISSEROW'S ARGUMENTS

Before proceeding to the specifics of their arguments, four clarificatory points are in order. First, I understand the kind of moral nihilism under investigation as being synonymous with the moral error theory, with the moral error theory defined as the view that all atomic, nontautological, positive, first-order moral judgments are assertoric yet untrue ("DSB," 7). Elliott and Isserow present moral realism as the alternative to moral nihilism but do not offer a definition of moral realism or a taxonomy of the kinds of metaethical views that fall under the umbrella of moral realism. But, as Elliott and Isserow note, distinguishing between different versions of moral nihilism and moral realism is not relevant for the task at hand because the ordinary human being's conceptions of metaethical views are unlikely to be especially fine-grained ("DSB," 8).²

Second, although Elliott and Isserow do not offer a definition of an "ordinary human being," they do, at different points, list a number of conditions that an ordinary human being would be likely to satisfy: an ordinary human being (henceforth "Alice") is likely to have ordinary human preferences and ordinary human beliefs ("DSB," 6), they are not a philosopher ("DSB," 8), and they care about a wide variety of things that cannot be reduced to their own positive feelings ("DSB," 10).

Third, the term '*wff*' is not to be taken too literally. As well as encapsulating the warm glow that Alice experiences when she acts rightly ("DSB," 22), the term '*wff*' can also encapsulate a sense of personal accomplishment, a sense of meaningful achievement ("DSB," 14), and the formation of a moral identity ("DSB," 23–25). Broadly, for the purposes of this article, the term '*wff*' can be treated as covering anything that Alice values that is tied to the satisfaction of her moral preferences.

Finally, Elliott and Isserow stipulate that whether Alice rationally ought to inquire after any given truth depends on whether doing so would be her subjectively rational choice, where her subjectively rational choice is that which will maximize her expected utility ("DSB," 10).

With those preliminaries out of the way, let us proceed with Elliott and Isserow's argument. After providing two toy cases that, it is argued, illustrate cases in which Alice rationally ought not to inquire after certain

2. Metaethical views that are neither forms of moral nihilism nor forms of moral realism, such as moral noncognitivism, are not discussed by Elliott and Isserow and can be set aside for the purposes of this article.

truths (“DSB,” 10–14), Elliott and Isserow provide a toy case designed to illustrate that Alice rationally ought not to inquire after the truth of moral nihilism:

Alice is deciding whether to help Bob, who is moving a sofa up a flight of stairs. On the one hand, Alice has no intrinsic desire to carry sofas upstairs, and all else equal she would prefer not to. On the other hand, there are several considerations in favor of helping.

First, Alice desires to help Bob because she cares about doing the right thing (whatever that may be), and she believes in this case that helping Bob is the right thing to do. Furthermore, whenever she does what she believes is the right thing, Alice gets a little warm fuzzy feeling inside. Alice enjoys this feeling, though it is by no means a primary driving force in favor of her doing the right thing generally. Over and above those considerations, Alice also desires to help Bob regardless of whether it is the right thing to do, simply because Bob is her friend and she wants to help her friends, and she also wants to avoid any social reprobation that might arise if it were to become widely believed that she is unhelpful.

As she is making up her mind, the oracle once again appears and offers to tell Alice whether moral nihilism is true, free of charge. Alice is open to the idea of nihilism—specifically, she’d assign it about 10 percent confidence—but the rest of her confidence resides in some form of moral realism. Should Alice accept? (“DSB,” 14)

Elliott and Isserow model Alice’s decision in the decision tree shown in figure 1. The outcomes are symbolized as follows:

- $x = \textit{Help}$ at a world where Realism is true, and Alice believes Realism.
- $y = \textit{Help}$ at a world where Nihilism is true, and Alice believes Realism.
- $z = \textit{Help}$ at a world where Nihilism is true, and Alice believes Nihilism.
- $q = \textit{Don't Help}$ at a world where Realism is true, and Alice believes Realism.
- $r = \textit{Don't Help}$ at a world where Nihilism is true, and Alice believes Realism.
- $s = \textit{Don't Help}$ at a world where Nihilism is true, and Alice believes Nihilism. (“DSB,” 15–16)

C symbolizes Alice’s credence in the relevant metaethical theory. Elliott and Isserow assume, for the sake of simplicity, that Alice believes that p if and only if $C(p) \geq 0.9$ (“DSB,” 15). Further, Elliott and Isserow assume that helping or not helping when Realism is true and $C(\textit{Realism}) = 0.9$ generates the same amount of utility as helping or not helping when Realism is true and $C(\textit{Realism}) = 1$ (“DSB,” 16).

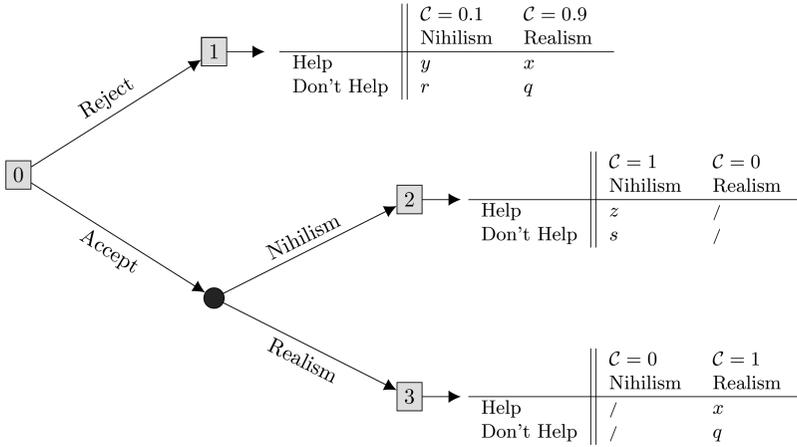


FIG. 1.—The Sofa (“DSB,” fig. 3.).

Finally, Elliott and Isserow posit the following three claims:

- A1.** At node 1, Help has maximum utility.
- A2.** x is at least as great as q .
- A3.** y is greater than either z or s . (“DSB,” 16)

Given A1, A2, and A3, no matter how the utility values contained within the decision tree are filled out, Reject will have greater expected utility. Therefore, given A1, A2, and A3, Alice rationally ought to reject the oracle’s offer to learn the truth of moral nihilism.

In support of A1, A2, and A3, Elliott and Isserow rely on the following three empirical assumptions:

CORRELATION. There is a correlation between Alice’s moral preferences and her nonmoral preferences, in the sense that she would usually prefer to do what she believes is the right thing regardless of the truth of nihilism.

COST. In worlds where nihilism is true but she believes it’s false, Alice still gets a pleasant *wff* for having done what she believes is the right thing—which she would not have if she came to believe there is no right thing to do.

NO COMPENSATION. The aforementioned cost of losing the *wff* is greater than any increase in utility to Help or Don’t Help (whichever is the greater) that results from coming to believe in nihilism at a world where it’s true. (“DSB,” 17)

If Alice would usually prefer to do what she believes is the right thing regardless of the truth of Nihilism (Correlation), then Help is Alice’s

subjectively rational choice at node 1 (A1), and Help will provide Alice with at least as much expected utility as Don't Help when Realism is true and she believes Realism to be true (A2). Hence, Correlation directly supports claims A1 and A2. Further, if Alice is precluded from experiencing *wff*'s when she believes Nihilism to be true (Cost), and if the cost to Alice's expected utility of being precluded from experiencing *wff*'s is not outweighed by any compensatory benefits (No Compensation), and given that Alice would usually prefer to do what she believes is the right thing regardless of the truth of Nihilism (Correlation), then even when Nihilism is true, Help when Alice believes Realism to be true will provide Alice with more expected utility than either Help or Don't Help when Alice believes Nihilism to be true (A3). Hence, given Correlation, Cost and No Compensation jointly support claim A3.

In the following section, I will be concerned only with rejecting Cost. In Section IV, I will proceed to show how rejecting Cost undermines the case for A3, and therefore how rejecting Cost threatens to undermine Elliott and Isserow's thesis that Alice rationally ought to reject the oracle's offer. I will proceed in Section IV to argue that Elliott and Isserow's defense of Correlation leaves them particularly vulnerable to attacks on Cost, and I present a dilemma that this poses for their view.

III. COST

Elliott and Isserow distinguish between Alice's *de dicto* and *de re* moral preferences: Alice *de dicto* prefers to do whatever is right and *de re* prefers, for each thing that is in fact morally right, to do that thing ("DSB," 21). After coming to learn the truth of moral nihilism, Alice would be unable to satisfy her *de dicto* moral preferences, for she would not believe anything to be morally right. But, given Correlation, Alice has a number of *de re* moral preferences that she does not stand to lose after having learned the truth of moral nihilism—"desires to help the global poor, promote peace in the Middle East, or save the whales, for example" ("DSB," 22).³ Elliott and Isserow provide two reasons to suspect that Alice would continue to hold on to her *de re* moral preferences after coming to learn the truth of moral nihilism: biological and cultural factors have selected for other-regarding emotional responses such as sympathy and guilt, while the prospect of

3. Strictly speaking, Alice's *de re* moral preferences would be best described as something other than "moral" preferences after coming to learn the truth of moral nihilism—perhaps "other-regarding preferences" more generally. Hereafter, given that Correlation predicts that the *de re* moral preferences that Alice has when she believes moral realism to be true are likely to be subsumed within the more general other-regarding preferences that Alice has after having come to learn the truth of moral nihilism, I will follow Elliott and Isserow in using the term '*de re* moral preferences' to cover both Alice's *de re* moral preferences when she believes moral realism to be true and her other-regarding preferences more generally when she has come to learn the truth of moral nihilism.

social disapproval provides Alice with indirectly self-interested reasons to calibrate her nonmoral preferences with moral considerations (“DSB,” 19–20). In order to defend Cost, Elliott and Isserow must therefore demonstrate that there are at least some *wff*’s tied to the satisfaction of Alice’s *de dicto* rather than *de re* moral preferences.

Elliott and Isserow’s argumentative strategy involves identifying two different kinds of *wff* and, for each kind of *wff*, providing one piece of evidence that demonstrates that the kind of *wff* under examination is tied to the satisfaction of Alice’s *de dicto* moral preferences. In this section, I will take each of the two kinds of *wff* in turn, and for each kind of *wff* I will (i) argue that the evidence provided by Elliott and Isserow undersupports their claim that the *wff* is tied to the satisfaction of Alice’s *de dicto* moral preferences and (ii) provide a positive argument as to why the *wff* in question can plausibly be tied to the satisfaction of Alice’s *de re* moral preferences. If the *wff*’s under investigation are in fact tied to the satisfaction of Alice’s *de re* moral preferences, then she can continue to experience those *wff*’s after coming to learn the truth of moral nihilism, and learning the truth of moral nihilism need not incur a cost for Alice.

The first kind of *wff* under examination regards the value that Alice places on her moral identity. According to Elliott and Isserow, it is “a fact now widely recognized in psychology that people care deeply about their moral selves” (“DSB,” 23), such that Alice’s self-conception as a morally good person forms an important part of her identity. But whether or not Alice can retain her moral self-conception after coming to learn the truth of moral nihilism depends on the factors that constitute Alice’s moral identity. If Alice’s moral identity has moral content—if she desires to be a person who does the morally right thing *de dicto*—then Alice would stand to lose her moral identity after coming to learn the truth of moral nihilism, for she could never satisfy her desire to do the morally right thing. On the other hand, if Alice’s moral identity comprises a set of *de re* moral preferences—if she desires to be helpful, to be honest, and to promote happiness—then Alice can retain her moral identity after coming to learn the truth of moral nihilism, for she can continue to hold and satisfy those *de re* moral preferences.

Elliott and Isserow point toward the Muhammad Ali effect as evidence that the moral commitment that underpins Alice’s moral identity is a *de dicto* moral preference to do the right thing (“DSB,” 23–25). The Muhammad Ali effect is a cognitive bias whereby people overestimate themselves in the moral domain more so than they overestimate themselves in other domains (such as that of intelligence). Elliott and Isserow highlight an explanation for the Muhammad Ali effect on which they claim that the evidence has converged: it is easier to verify other qualities, such as intelligence, than it is to verify the strength of someone’s moral character. This “interpretational or attributional ambiguity” renders it easier for people to get away with convincing themselves and others that

they have a stronger moral character than it is to convince themselves and others that they are especially intelligent.⁴ But, argue Elliott and Isserow, this explanation for the Muhammad Ali effect would not have legs if the important part of one's moral identity lacked *de dicto* moral content—for example, if people merely desired to be seen as helpful or honest, for such qualities are much more easily verifiable than the ambiguous quality of being a good person. The upshot for Elliott and Isserow is that if the important part of Alice's moral identity contains *de dicto* rather than *de re* moral preferences, then Alice would stand to lose her moral identity after coming to learn the truth of moral nihilism.

Elliott and Isserow cite studies by Scott Allison, George Goethals, and David Messick and by Paul Van Lange as evidence for their claim that the Muhammad Ali effect is explained by the nonverifiability of one's moral character.⁵ However, these studies are most plausibly read as favoring an alternative explanation for the Muhammad Ali effect: the desirability of having a strong moral character. The first study, conducted by Allison et al. in 1989, found that subjects were more likely to believe themselves to frequently perform moral behaviors than they were to believe themselves to frequently perform intelligent behaviors and dubbed this effect the Muhammad Ali effect. Though Allison et al. did not test for particular explanations for the Muhammad Ali effect, they speculated that egocentric biases are likely to be more frequent among attributes that are less public, specific, and objective and that morality is less public, specific, and objective than intelligence.⁶ The second study, conducted by Van Lange in 1991, considered three candidate explanations for the Muhammad Ali effect: people might overestimate their moral character more so than intelligence because they judge morality to be more positive an attribute than intelligence (desirability), because they judge morality to be more controllable than intelligence (controllability), or because they judge morality to be less easily verifiable than intelligence (verifiability). Subjects were asked to recount examples of their own behavior and other's behavior and to rate the behaviors as more or less moral or intelligent. Independent judges rated the extent to which the behaviors described were positive, controllable, and verifiable and found that subjects described their own behavior as more positive and more controllable, but there was no difference in verifiability across descriptions of their own and other's behavior.⁷ Van Lange also

4. Paul A. M. Van Lange, "Being Better but Not Smarter Than Others: The Muhammad Ali Effect at Work in Interpersonal Situations," *Personality and Social Psychology Bulletin* 17 (1991): 689–93, 692.

5. Scott T. Allison, George R. Goethals, and David M. Messick, "On Being Better but Not Smarter Than Others: The Muhammad Ali Effect," *Social Cognition* 7 (1989): 275–95; Van Lange, "Being Better but Not Smarter."

6. Allison, Goethals, and Messick, "On Being Better," 289–94.

7. Van Lange, "Being Better but Not Smarter," 691–92.

found that the Muhammad Ali effect was stronger when the behavior described was especially positive than when it was especially controllable, and he concluded that “differences in the positivity of the behaviors that subjects described for self and others account, at least in part, for the Muhammad Ali effect observed in the present study.”⁸ In a later study with Constantine Sedikides, Van Lange tested for each of the three candidate explanations for the Muhammad Ali effect (desirability, controllability, and verifiability).⁹ Van Lange and Sedikides found again that judgments about the desirability of the attribute under investigation mediated the Muhammad Ali effect while judgments about controllability and verifiability did not mediate the Muhammad Ali effect. Van Lange and Sedikides concluded that “the *actual* mechanism underlying the Muhammad Ali effect, as observed in the present study, would seem to be rooted in the tendency to regard honesty as more desirable than intelligence.”¹⁰

The desirability explanation for the Muhammad Ali effect threatens to undermine Elliott and Isserow’s argument that Alice’s moral identity is underpinned by her *de dicto* rather than her *de re* moral preferences. If the Muhammad Ali effect is explained by the desirability rather than the nonverifiability of having a strong moral character, there is no reason to suppose that judgments about moral character are especially nonverifiable and no reason to suppose that Alice’s moral identity cannot be underpinned by *de re* moral preferences. Further, if Alice’s moral identity can be underpinned by *de re* moral preferences, there is no reason to suppose that Alice would lose her moral identity after coming to learn the truth of moral nihilism—Alice can still conceive of herself as helpful, kind, and someone who promotes happiness.¹¹

We may be able to push this point further: not only is it consistent with the desirability explanation for the Muhammad Ali effect that Alice can retain her moral identity after coming to learn the truth of moral nihilism, but the desirability explanation might even provide positive reason to suspect that Alice can retain her moral identity after coming to learn the truth of moral nihilism. To explain: if the desirability of having a strong moral character gives rise to the Muhammad Ali effect, this means that having

8. *Ibid.*, 692.

9. Paul A. M. Van Lange and Constantine Sedikides, “Being More Honest but Not Necessarily More Intelligent Than Others: Generality and Explanations for the Muhammad Ali Effect,” *European Journal of Social Psychology* 28 (1998): 675–80.

10. *Ibid.*, 680. Honesty was used as an indicator for morality in the study.

11. In n. 3 I noted that Alice’s *de re* moral preferences would be best described as something other than “moral” preferences after coming to learn the truth of moral nihilism—perhaps “other-regarding preferences” more generally. Likewise, if Alice’s moral identity is underpinned by *de re* moral preferences, and if Alice can retain her moral identity after coming to learn the truth of moral nihilism, her moral identity would be best described as something other than a “moral” identity after coming to learn the truth of moral nihilism—perhaps a “prosocial identity” more generally.

a strong moral character is sufficiently desirable that it moves people to perform a kind of mental gymnastics whereby they perceive themselves to be more moral than their behavior would indicate. If the desirability of having a strong moral character moves people to perform that kind of mental gymnastics, then perhaps it could also move Alice to perform a kind of mental gymnastics whereby she believes herself to have strong moral credentials even after coming to learn the truth of moral nihilism. This kind of mental gymnastics need not be especially convoluted if Alice were to ground her moral self-conception in her *de re* moral preferences—Alice’s moral self-conception could be as someone who is helpful, is honest, and promotes happiness, rather than as someone who does the right thing *de dicto*. Granted, Alice’s moral identity would lack genuinely moral content, but insofar as Alice cares deeply about satisfying her *de re* moral preferences (and Elliott and Isserow grant that she does; “DSB,” 22), it is not clear that anything important would be lost.

The second kind of *wff* identified by Elliott and Isserow is the positive experiential effect that human beings undergo when their moral preferences are satisfied (“DSB,” 25–26). As with Alice’s moral identity, Elliott and Isserow argue that there are positive experiential effects tied to the satisfaction of *de dicto* moral preferences that Alice would stand to lose after coming to learn the truth of moral nihilism. Elliott and Isserow offer as evidence the role of moral judgments in viewers’ enjoyment of dramatic entertainment: viewers tend to find dramas more enjoyable when characters get their just deserts.¹² Children offer a particularly pronounced example—children around the age of four tend to prefer greater amounts of retribution, while by the age of eight children develop a sense of justice according to which they prefer proportionate retaliation. These preferences manifest in their enjoyment of fairy tales: children around the age of four prefer fairy tales that display greater levels of retribution, while children around the age of eight prefer fairy tales that display proportionate retaliation.¹³

However, it is unclear why Elliott and Isserow take this evidence to point toward the role of satisfied *de dicto* rather than *de re* moral preferences in the enjoyment of dramatic entertainment. In the single study cited relating to fairy tales, it is consistent with Zillman and Bryant’s findings that the positive experiential effect arises from the children’s *de re* moral preference for excessive retaliation (in the case of four-year-olds)

12. Arthur A. Raney and Jennings Bryant, “Moral Judgment and Crime Drama: An Integrated Theory of Enjoyment,” *Journal of Communication* 52 (2002): 402–15; Arthur A. Raney, “Moral Judgment as a Predictor of Enjoyment of Crime Drama,” *Media Psychology* 4 (2002): 305–22; Arthur A. Raney, “Punishing Media Criminals and Moral Judgment: The Impact on Enjoyment,” *Media Psychology* 7 (2005): 145–63.

13. Dolf Zillman and Jennings Bryant, “Viewer’s Moral Sanction of Retribution in the Appreciation of Dramatic Presentations,” *Journal of Experimental Social Psychology* 11 (1975): 572–82.

or for proportionate retaliation (in the case of eight-year-olds) rather than from the satisfaction of their *de dicto* moral preference for the right outcome. In the case of dramas, the three studies cited by Elliott and Isserow purport to demonstrate that satisfied moral judgments predict subjects' enjoyment of dramas.¹⁴ Raney and Bryant used subjects' judgments about the deservedness of punishment as indicators for moral judgments, while Raney in his two solo-authored studies used subjects' judgments about the deservedness of punishment, as well as their self-reported feelings of sympathy toward the victim, as indicators for moral judgment. It is consistent with each of the three studies that the positive experiential effect arises from satisfied *de re* moral preferences for punishment toward those who have harmed characters that the subject feels sympathetic toward rather than from satisfied *de dicto* moral preferences. Finally, Elliott and Isserow also cite an essay by Dolf Zillman and Joanne Cantor that develops a "disposition theory" according to which the intensity of a respondent's appreciation of a humorous presentation is heightened when the respondent is antipathetic toward disparaged protagonists and impaired when the respondent is sympathetic toward the disparaged protagonist.¹⁵ But it is consistent with Zillman and Cantor's disposition theory that the positive experiential effect arises from satisfied *de re* moral preferences for disliked protagonists to be disparaged and for liked protagonists to be rewarded rather than from satisfied *de dicto* moral preferences. Correlation predicts that Alice, even after coming to learn the truth of moral nihilism, would continue to hold the *de re* moral preferences that plausibly give rise to the positive experiential effect in each of these five studies: Elliott and Isserow explicitly argue that Alice would continue to experience feelings of sympathy in response to suffering and feelings of anger that fuel punitive action ("DSB," 19). Therefore, in each of the five studies cited, it is consistent with the evidence that Alice, even after coming to learn the truth of moral nihilism, could continue to undergo the positive experiential effect.

By their own admission, Elliott and Isserow face a disentanglement problem: given that satisfied *de dicto* moral preferences tend to be accompanied by satisfied *de re* moral preferences, it is difficult to determine which preference gives rise to the *wff*. In the absence of empirical data on the effects of coming to learn the truth of moral nihilism, an illuminating strategy would be to identify positive experiential effects that arise amid a disharmony between *de dicto* and *de re* moral preferences. Such conditions do not frequently obtain, but one study conducted by Daniel Västfjäll et al.

14. Raney and Bryant, "Moral Judgment and Crime Drama"; Raney, "Moral Judgment as a Predictor"; Raney, "Punishing Media Criminals."

15. Dolf Zillman and Joanne Cantor, "A Disposition Theory of Humour and Mirth," in *Humor and Laughter: Theory, Research, and Application*, ed. Anthony J. Chapman and Hugh C. Foot (London: Wiley, 1976), 93–115.

gestures in this direction. Västfjäll et al. found a singularity effect, related to the identifiable victim effect, whereby the positive feelings experienced by subjects, evidenced by self-reports and physiological indicators, were at their strongest when they gave money to help one needy child and subsequently diminished with each additional child helped.¹⁶ It is highly unlikely that the subjects were more inclined to judge themselves to have been doing the right thing *de dicto* when providing assistance to the individual child than when providing assistance to multiple children. Much more plausible is the idea that there are other factors that gave rise to the subjects' positive experiential effects when they helped the single child, such as an increased capacity to feel sympathy with one person than with multiple persons.

This study is of particular significance for the case of Alice because it highlights an instance not only in which there are other factors besides *de dicto* moral preferences that give rise to *wff*'s but also in which the betterness of an outcome *de dicto* is correlated with diminishing levels of *wff*'s. Although the singularity effect does not rule out the possibility that *de dicto* moral preferences might be one factor among others in giving rise to positive experiential effects, it does suggest that the capacity to experience other-regarding sentiments such as sympathy can also play a significant role in giving rise to positive experiential effects. Given that, as Elliott and Isserow concede, moral nihilists are likely to retain these other-regarding sentiments after coming to learn the truth of moral nihilism ("DSB," 22), Alice may well continue to experience many of these positive experiential effects after having come to learn the truth of moral nihilism.

IV. A DILEMMA

In Section III, I argued that the evidence provided by Elliott and Isserow does not give reason to think that there are *wff*'s tied to the satisfaction of Alice's *de dicto* moral preferences, and I gave independent evidence to suggest that Alice's *wff*'s are in fact tied to the satisfaction of her *de re* moral preferences. More evidence on the psychological effects of believing moral nihilism is needed to adjudicate between proponents and opponents of Cost. But even if my arguments in Section III are not strong enough to warrant the rejection of Cost, I hope to have provided sufficient reason to doubt that Elliott and Isserow have successfully defended Cost.

If Cost is not vindicated, this poses a grave challenge for Elliott and Isserow. We saw in Section II that Cost is needed to establish A3, and without A3, the utility values in figure 1 can be filled out in such a way

16. Daniel Västfjäll et al., "Compassion Fade: Affect and Charity Are Greatest for a Single Child in Need," *PLoS One* 9 (2014): e100115.

that Accept has greater expected utility than Reject. If the utility values can be filled out in such a way that Accept has the greater expected utility, then it is possible that Alice rationally ought to accept the oracle's offer, and it is possible that it is in Alice's best interests to learn the truth of moral nihilism.¹⁷

The task of defending Cost may have proven easier for Elliott and Isserow had they not defended Correlation: it is difficult to argue at once that Alice would retain her *de re* moral preferences after coming to learn the truth of moral nihilism and that Alice would miss out on important *wff*'s after coming to learn the truth of moral nihilism, because those *wff*'s can plausibly arise from the satisfaction of Alice's *de re* moral preferences. In fact, my arguments in Section III regarding the two different kinds of *wff* under examination depend on the truth of Correlation—in particular, the claim that Alice would retain her *de re* moral preferences after coming to learn the truth of moral nihilism. With regard to Alice's moral identity, I argued that Alice could retain her moral self-conception if she grounded her moral self-conception on her *de re* moral preferences, and with regard to Alice's positive experiential effects, I argued that the satisfaction of Alice's *de re* moral preferences can give rise to her positive experiential effects. Each of these avenues of response would have been blocked had Elliott and Isserow rejected Correlation and argued that Alice would lose her *de re* moral preferences after coming to learn the truth of moral nihilism. Note that this tack is not without precedent: Guy Kahane, for example, has argued that one would likely suffer a wholesale loss of one's subjective concerns after coming to believe in nihilism.¹⁸

17. I have argued that Elliott and Isserow undersupport the claim that Alice would suffer costs from learning the truth of moral nihilism. However, I have not argued that Alice would enjoy benefits from learning the truth of moral nihilism. In the absence of arguments to the effect that Alice would indeed enjoy benefits from learning the truth of moral nihilism, Elliott and Isserow could fall back on a weakened version of A3, A3*, formulated thus: "y is at least as great as z and s." If A3 were replaced with A3*, the utility values in fig. 1 could be filled out in such a way that the expected utility of Accept is as great as, though no greater than, Reject. But to vindicate A3* without vindicating Cost would require defending a stronger version of No Compensation according to which Alice would enjoy no benefits from learning the truth of moral nihilism whatsoever. Not even Elliott and Isserow defend this strong claim; they concede that Alice may enjoy benefits from learning the truth of moral nihilism (e.g., freedom from moral guilt and the value of true belief), though they argue that these benefits are "extremely miniscule" ("DSB," 30) and likely to be outweighed by the costs of learning the truth of moral nihilism ("DSB," 28–31). In any case, even if Elliott and Isserow were to replace A3 with A3*, the expected utility of Accept could still be as great as the expected utility of Reject, Accept need not be subjectively irrational, and their thesis that ordinary human beings rationally ought not to inquire after the truth of moral nihilism would not go through.

18. Guy Kahane, "If Nothing Matters," *Noûs* 51 (2017): 327–53.

However, had Elliott and Isserow rejected Correlation, they would have faced at least three separate challenges. Perhaps most pressingly, Correlation is the empirical assumption employed in order to support claims A1 and A2, and without A1 and A2, the utility values in figure 1 can be filled out in such a way that Accept has the greater expected utility. If the utility values can be filled out in such a way that Accept has the greater expected utility, then it is possible that Alice rationally ought to accept the oracle's offer, and it is possible that it is in Alice's best interests to learn the truth of moral nihilism.

Secondly, defending Correlation helps stabilize No Compensation. If Correlation is true and Alice's nonmoral preferences correlate with her moral preferences, then coming to learn the truth of moral nihilism would not significantly alter the kinds of actions that Alice would prefer to perform, so the compensatory benefits of coming to learn the truth of moral nihilism are unlikely to be especially large. By contrast, if Correlation were not true and Alice would prefer to perform radically different kinds of actions after coming to learn the truth of moral nihilism, there would be an increased likelihood that Alice's utility after coming to learn the truth of moral nihilism would diverge from her utility prior to coming to learn the truth of moral nihilism, and the thesis that Alice would experience no significant compensatory benefits from coming to learn the truth of moral nihilism would be less clear.

Finally, efforts to reject Correlation would raise the immediate problem of overcoming the evidence and argumentation that Elliott and Isserow offer in favor of Correlation. Elliott and Isserow would be tasked, for example, with explaining why neither Alice's capacities for sympathy and guilt nor the prospect of social disapproval suffice to calibrate her nonmoral preferences with moral considerations. Further, Elliott and Isserow would also be tasked with explaining how the nontrivial number of philosophers who subscribe to variants of moral nihilism appear to have avoided wholesale losses of their *de re* moral preferences.

These remarks pose a dilemma for Elliott and Isserow. If, on the one hand, Elliott and Isserow defend Correlation, then, as I have argued, their task of defending Cost is made much more difficult, and if they cannot defend Cost, then the utility values in figure 1 can be filled out such that it is in Alice's best interests to learn the truth of moral nihilism. On the other hand, if Elliott and Isserow reject Correlation, then once again the utility values in figure 1 can be filled out in such a way that it is in Alice's best interests to learn the truth of moral nihilism, while this would also risk destabilizing their defense of No Compensation and would incur the further difficulty of explaining away the evidence that they provided in favor of Correlation. Proponents of the view that one rationally ought not to inquire after the truth of moral nihilism must overcome either one of these sets of challenges.

V. CONCLUSION

Whether or not Alice rationally ought to accept the oracle's offer to learn the truth of moral nihilism hinges on whether (and if so, to what extent) she will face costs that arise from coming to learn the truth of moral nihilism and whether (and if so, to what extent) she will enjoy compensatory benefits that arise from coming to learn the truth of moral nihilism. I do not intend to have argued conclusively in either direction, and I suspect that further empirical evidence may be required in order to make headway in revealing the likely consequences of belief in moral nihilism. The more modest aims of this article that I hope to have fulfilled are to provide reasons to doubt that coming to learn the truth of moral nihilism would incur a cost for ordinary human beings, and consequently to provide reasons to doubt that an ordinary human being rationally ought not to inquire after the truth of moral nihilism, and to outline a dilemma that must be overcome by proponents of the view that ordinary human beings rationally ought not to inquire after the truth of moral nihilism.