Don't Stop Believing (Hold onto That Warm Fuzzy Feeling)*

Edward J. R. Elliott and Jessica Isserow

If beliefs are a map by which we steer, then, ceteris paribus, we should want a more accurate map. However, the world could be structured so as to punish learning with respect to certain topics—by learning new information, one's situation could be worse than it otherwise would have been. We investigate whether the world is structured so as to punish learning specifically about moral nihilism. We ask, if an ordinary person had the option to learn the truth about moral nihilism, ought she to take it? We argue, given plausible assumptions about ordinary human preferences, she (probably) should not.

I. INTRODUCTION

All else being equal, it's good to have true beliefs. On a common view of human behavior and decision-making, we are for the most part pragmatically rational beings: we typically act in such a way as to bring about the kinds of things we want, given the ways we take the world to be. To borrow a metaphor from Ramsey, our beliefs are a map by which we steer our efforts to

* Both authors contributed substantially to this work; order of authorship was determined by the flip of a fair coin. Thanks are due to Heather Browning, Toby Handfield, Michael Huemer, Norbit Paulo, Kim Sterelny, Pekka Väyrynen, Jack Woods, and audiences at the University of Sydney, the 2018 Evolution & Epistemology conference (Utrecht), a Leeds CMM workshop, the ACU 2018 international moral epistemology conference (Melbourne), and the 2018 AAP/AAPNZ (Wellington). We apologize for any omissions. We are grateful to those referees and associate editors who made helpful and constructive suggestions for the article.

Ethics, volume 132, number 1, October 2021.
 https://doi.org/10.1086/715291

bring the world more in line with the way we'd prefer it to be.[1] If this is so, then it stands to reason that a more accurate map will usually be better than one that says (for example) that there's a road where no road exists, a forest where there are no trees, or a mountain where there's only a molehill.

As a rough-and-ready generalization, then: the more accurate an agent's beliefs are, the more likely it will be that the subjectively rational choice (that which would maximize her preference satisfaction if her beliefs were fully accurate) and the objectively correct choice (that which would actually maximize her preference satisfaction) will coincide for any decision situation that she might find herself in. Indeed, in the limiting case where an agent—we will call her Alice—has complete knowledge of what her world is like and where she's situated within it, she will typically choose whatever action is available to her that will result in what is by her lights the best outcome. That is, if Alice had only true beliefs and were ignorant of nothing relevant to her present decision, then the rational choice and the correct choice for her would be one and the same.

With these considerations in mind, it is plausible that if some philosophical theory is true, we should usually want to learn of its truth—especially inasmuch as the truth or falsity of that theory could have far-reaching implications as to whether and how well our preferences are satisfied. And for many of us, *moral nihilism* is just such a theory. Briefly—we'll precisify below—the moral nihilist says that there are no moral facts: no facts about who is morally good or bad, about what is morally right or wrong, or about what one morally ought or ought not to do. Morality, in general, is bunk. If such a theory were true, then this would make a difference to how we evaluate the potential consequences of our actions. Most of us care about being morally good agents and about choosing the morally right action, and if it turns out that there are no morally good people, or no morally right actions to choose, then, ceteris paribus, this seems the sort of thing we should want to know about as soon as possible.

However, we should first check that the *ceteri* really are *paribu*.[2] After all, it's easy to see that it's not necessarily true that we always do better by improving the accuracy of our beliefs whenever we're given the choice to do so. Consider the evil demon who hates know-it-alls: the more that Alice learns about the world around her, the more the demon limits Alice's options to only those with the worst outcomes. In the limiting case, Alice knows exactly which of her choices will maximize her preference satisfaction given whatever decision situation she finds herself in, so she always makes the best choices she can—but by virtue of her now perfectly ideal

1. Frank P. Ramsey, "General Propositions and Causality," in *The Foundations of Mathematics and Other Logical Essays*, ed. R. B. Braithwaite (London: Austin, 1931), 237–55.

2. Only one of the authors understands Latin, and that author had no part in writing this sentence.

epistemic state, any decision situation she's in will be much worse than it might have been otherwise.

So it is at least possible for the world to be structured so as to punish general improvements to one's epistemic state. It may also be structured so as to punish improvements with respect to specific topics. If an overzealous moral realist credibly threatened to set off a nuclear weapon were Alice to learn any more about the truth of moral nihilism, then she may quite rationally decide to avoid any further inquiries on the topic.

The examples just given are fanciful, but they do raise an interesting question: just how plausible is it that the world the average person lives in is structured so as to punish learning specifically with respect to moral nihilism? Or to put that question in a slightly different way: if Alice, whom we will suppose henceforth is an ordinary human being with ordinary human preferences and ordinary human beliefs, had the option to learn about the truth of moral nihilism, totally free of charge, then ought she to take it?

We will argue that Alice would (probably) be irrational to take the offer, provided her preferences and beliefs conform to (what we will argue are) common and perfectly reasonable patterns. We do not put this conclusion forward as a necessary claim. It is, of course, possible in some circumstances to rationally choose to learn more about the truth of moral nihilism. Nor do we want to say that our conclusion applies to everyone alive today. People vary, some more so than others. In fact, we will argue that philosophers in particular can have incentives to inquire after the truth of moral nihilism that plausibly outweigh the costs we think are involved. But philosophers are unusual. Alice isn't a philosopher, and if she is the way we think she is—that is, the way most people are—then she would do better to avoid inquiry into moral nihilism.

The remainder of the article is structured as follows. In Section II we say a little more to pin down the sort of nihilistic theory we have in mind, and we lay out some general background assumptions that will be employed in the ensuing discussion. Then, in Section III we introduce a standard framework for thinking about the value of learning using some simple hypothetical cases, and in Section IV we apply that same framework to the case of learning about moral nihilism. Finally, in Sections V–VII we provide an empirical case for the key assumptions about beliefs and preferences that we require for our conclusion, and in Section VIII we discuss some objections and complications.

Unlike much previous work, which focuses primarily on the social costs of widespread belief in moral nihilism, our focus lies squarely with the costs to individuals. With that said, some readers may very well take our arguments to have social policy implications—to support a kind of "government house" moral nihilism. We will return to this matter briefly in the conclusion without dwelling on it, leaving the reader free to draw any social policy lessons for themselves. We also note that our arguments

concern the likely consequences of inquiring into moral nihilism. Similar arguments may or may not also apply to, say, the rational advisability of inquiry into animalism or modal realism. But such arguments would require the careful consideration of their own distinctive supporting evidence, so we do not take a stand on such matters here.

II. BACKGROUND ASSUMPTIONS

It will be helpful to begin by laying out some key ideas and assumptions that we will be making with regard to moral nihilism, as well as how we will be understanding beliefs, preferences, rational choice, and the relationships between them.

A. Moral Nihilism

The moral nihilist—she may also go by *moral error theorist*—is a cognitivist with respect to moral discourse, taking ordinary moral claims to be in the market for truth and falsity. However, she parts company from other cognitivists—so-called *success-theorists*—in taking such claims to be systematically false.[3] Alternatively, some among her ranks may take moral discourse to fall victim to presupposition failure and hence may prefer to characterize moral claims as neither true nor false.[4] The important point is that the moral nihilist systematically denies the truth of ordinary moral claims.

Some subtlety is called for here. The moral nihilist may very well be able to stomach the truth of some moral claims. She might allow that some nonatomic ('Either lying is wrong or the Eiffel Tower is in Paris'), tautological ('Wrongness is wrongness'), negative ('Stealing isn't wrong'), or second-order ('There are no moral facts') moral claims could still be true. But we take it that such exceptions will be of cold comfort to the opponent of moral nihilism. At the very least, the moral nihilist will want to say that all atomic, nontautological, positive, first-order moral claims are not true. This, we submit, is a sizable portion of moral discourse—sizable enough to render moral nihilism a prima facie unsettling proposal.

Different philosophers have had different grounds for endorsing moral nihilism. Perhaps moral facts would be unacceptably "queer," the sorts of things that could not hope to find a place in the natural world.[5] Or

3. For a helpful taxonomy, see Geoff Sayre-McCord, "The Many Moral Realism," *Southern Journal of Philosophy* 24 (1986): 1–22.

4. See Richard Joyce, *The Myth of Morality* (Cambridge: Cambridge University Press, 2001), chap. 1; and Wouter Floris Kalf, "Moral Error Theory, Entailment and Presupposition," *Ethical Theory and Moral Practice* 16 (2013): 923–37.

5. John L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin, 1977).

maybe such facts would be explanatorily idle, swiftly eliminated from one's ontology with an unforgiving swipe of Occam's razor.[6] Or perhaps our moral talk is underwritten by a problematic commitment to categorical reasons.[7]

This also means that there are a variety of moral nihilisms that have been put forward over the years—and we will be riding roughshod over the distinctions between them. But there is a reason for this! Whereas philosophers will want to draw nuanced distinctions between manifold varieties of metaethical theory, our question concerns Alice—whom we're assuming is not a philosopher. What matters for our arguments will be Alice's conceptions of moral realism and moral nihilism, and we cannot expect Alice to be aware of (or care about) the sometimes very subtle distinctions between realist and nihilist theories that we philosophers have spilled so much ink over.

In effect, then, you can take our use of 'moral nihilism' in what follows to designate the disjunction of more specific varieties of nihilism that philosophers have (or might) put forward. Some of those disjuncts will be what we can call *metaphysically necessary* theories—that is, varieties of moral nihilism such that if they're true they must be true as a matter of metaphysical necessity. And some of the disjuncts will be *epistemically necessary*—that is, theories such that if they're true they will be true a priori. Some might be both. But we take it that metaphysically contingent nihilist theories make sense as well, and (as far as we can tell) at least some of these are not decidable a priori.[8]

Finally, it also bears mentioning here that we conceive of moral nihilism as a local nihilism. That is to say, the kind of moral nihilism at issue is not merely a symptom of a more sweeping, global nihilism, according to which no normative claims are true.[9] We hasten to emphasize that this assumption is not idiosyncratic. Global normative nihilism is plausibly the exception rather than the rule.[10] In what follows, then, we will happily help ourselves to normative language—for instance, in speaking of how ordinary agents like Alice rationally ought to choose.

6. Jonas Olson, *Moral Error Theory: History, Critique, Defence* (New York: Oxford University Press, 2014), 123–36.

7. Joyce, *Myth of Morality*.

8. See Kristie Miller, "On Contingently Error-Theoretic Concepts," *American Philosophical Quarterly* 47 (2010): 181–90.

9. See Bart Streumer, *Unbelievable Errors: An Error Theory about All Normative Judgements* (New York: Oxford University Press, 2017).

10. See Richard Joyce and Simon Kirchin, introduction to *A World without Values: Essays on John Mackie's Moral Error Theory*, ed. Richard Joyce and Simon Kirchin (New York: Springer, 2010), ix–xxiv, xiii.

B. Beliefs, Preferences, and Decision-Making

So that's enough hedging about moral nihilism; let's now talk about our understanding of beliefs, preferences, and rational decision-making. For this we will be adopting a generally Bayesian approach.[11]

With respect to Alice's beliefs, this requires the existence of a credence function, C, which assigns numerical strengths of belief between 0 and 1 to the propositions regarding which Alice has opinions (or "is aware of") and eo ipso represents her beliefs in toto. We will assume that the following conditions on C are all at least roughly true:

1. If p is inconsistent, then $C(p) = 0$ and $C(\neg p) = 1$.
2. If $p_1, \ldots, p_n$ are mutually exclusive propositions of which Alice is aware, then $C(p_1 \vee \ldots \vee p_n) = C(p_1) + \ldots + C(p_n)$, at least for small n.
3. It's rationally permissible that $0 < C(\text{Nihilism}) < 1$.

The Bayesian approach also requires the existence of a utility function, U, which provides a numerical measure of the extent to which Alice's preferences are satisfied given different ways the world might be. We prefer (but do not require for our arguments) a Humean account of rational preference. That is, beyond basic coherence requirements like transitivity, there are no restrictions on what an agent's preferences ought to look like. A rational agent doesn't have to care, qua rational agent, about what's morally good, or "the truth," or indeed about anything else whatsoever. This also means that Alice's preferences need not depend essentially on matters of personal experience—that is, our use of 'utility' is not to be interpreted as a measure of some experiential state of pleasure, happiness, or a sense of satisfaction. In terms borrowed from the economics literature, we're interested in what's often called *decision utility*, rather than *experienced* or *hedonic utility*.[12] For example, suppose that Alice cares overwhelmingly about maximizing the number of puppies there are, and that world ω_1 has many more puppies than ω_2. Then, ω_1 carries more utility for Alice than

11. A referee notes that we don't need Bayesianism to make our argument, which could be made in a rough way without all of the formalisms. Perhaps this is true. But rough arguments lead to rough conclusions. Our argument rests on drawing comparisons between trade-offs among a decision-maker's choices that are sensitive specifically to the relationship between values assigned to the outcomes of those choices and the decision-maker's degrees of belief over relevant states of the world. We don't see how we could fruitfully draw such comparisons without doing so within a formal framework designed to handle exactly these kinds of trade-offs. Moreover, by making use of the Bayesian framework, we are making the structure of our argument and our assumptions as explicit as we reasonably can, so that you, the reader, can know exactly where it is you want to disagree if you so choose.

12. For more on these distinctions, see Roberto Fumagalli, "The Futile Search for True Utility," *Economics and Philosophy* 29 (2013): 325–47.

ω_2 does—even if at ω_2 Alice has more puppies than she knows what to do with, while at ω_1 Alice sadly believes that puppies have gone extinct, and she never gets to experience the joys of having any puppies around for herself. Preference satisfaction doesn't imply awareness of that satisfaction, and utility isn't a measure of how nice it feels when you believe your preferences have been satisfied. (Most ordinary people usually do care about having nice feelings, of course, and this fact will be important for our argument—but most ordinary people tend to care about a lot more besides just having nice feelings, and that, too, will matter for what we have to say.)

Finally, we assume that the subjectively rational choice in any decision situation is that which maximizes expected utility. Where Alice has to choose among some collection of options, each of which has a different outcome under different ways the world might be consistent with what she believes, Alice should pick the option (or one of the options) with the greatest *C*-weighted average utility as measured by *U*. In more formal terms, supposing that if

a) $\{p_1, \ldots, p_n\}$ is a finite partition of propositions, where each element is causally and evidentially independent of whatever option Alice decides on, and

b) option A has outcome a_i if p_i is true, and B has outcome b_i if p_i is true $(i = 1, \ldots, n)$,

then Alice should weakly prefer option A to B if and only if

$$\sum_{i=1}^{n} C(p_i)U(a_i) \geq \sum_{i=1}^{n} C(p_i)U(b_i).$$

We will not defend this account of rational choice here. That's been done more than enough elsewhere. We do note, however, that alternative accounts of rational choice in which risk aversion plays a bigger role will tend to favor our conclusions, given our empirical assumptions.

III. THE VALUE OF LEARNING

With all that out of the way, how should we think about the value of new information within the general framework we've outlined? We will introduce this with some hypothetical cases.

> **Case 1: The Bet.** Before Alice and Bob sits an opaque box, which contains either a red ball or a blue ball. Alice doesn't know what color the ball is, but she's slightly more confident that it's red—specifically, $C(\text{Red}) = 0.6$. Alice knows that Bob doesn't know what color the ball is. Bob offers Alice a bet: he'll reach into the box and pull out the ball; if it's red, then Alice wins \$10, and if it's blue, then Alice pays him \$10.

> As Alice is deciding whether to take the bet, an oracle appears. The oracle is always perfectly reliable, and Alice knows this. She offers to tell Alice whether the ball is red or blue before she decides. The offer is free of charge, and there are no strings attached. Alice knows she can trust the oracle. Should she accept?

It is immediately obvious that Alice should accept. But we can slow things down and rationally reconstruct her reasoning using figure 1.

At node 0, Alice is trying to decide whether to *Reject* or *Accept* the oracle's offer. If she decides to *Reject*, then she knows she will end up in the decision situation at node 1, where she has to decide between *Bet* and *Don't Bet* given her unchanged belief state. Letting utilities equal the dollar values for simplicity, in this case the expected utility at node 1 (written EU_1) of *Bet* is $EU_1(Bet) = (0.6 \times 10) + (0.4 \times -10) = 2$, which is higher than the expected utility of *Don't Bet*. So, Alice reasons that if she were in that situation, she'd certainly choose *Bet*; hence, the expected utility of *Reject* at node 0 is just the expected utility of the choice she'd end up making if she rejected—that is, $EU_0(Reject) = EU_1(Bet)$. On the other hand, if she decides to *Accept*, then she knows that the oracle is going to tell her either that the ball is *Red* or that it's *Blue*. She doesn't know what she'll be told, but she does have beliefs about which is more likely. She reasons that there's a 60 percent probability that she'll be told the ball is red, in which case she'll be in the decision situation at node 2, where she'd choose *Bet* and will win a guaranteed \$10. On the other hand, there's a 40 percent probability that the oracle will tell her the ball is blue, in which case she'll

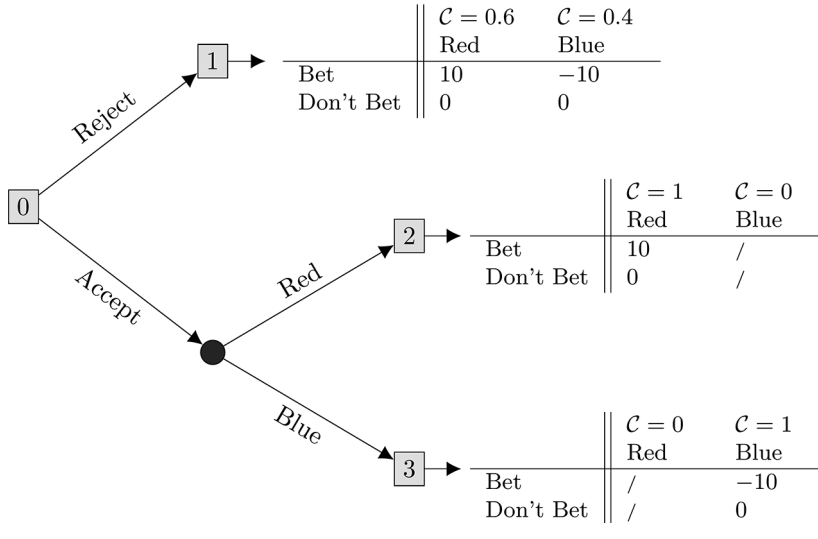

Fig. 1.—The Bet.

be in the decision situation at node 3, whereupon she'd want to avoid a sure loss and choose *Don't Bet*. Overall, then, the expected utility of *Accept* at node 0 is $[0.6 \times EU_2(Bet)] + [0.4 \times EU_3(Don't\ Bet)] = (0.6 \times 10) + (0.4 \times 0) = 6$, which is greater than that of *Reject*. So Alice takes up the oracle's offer.

Note a special feature of the case: accepting the oracle's offer comes with no associated costs. Alice doesn't have to pay any money for the information, there's no cost in time or effort, she doesn't have to promise away her first-born child, and so on. Nor is Alice required to forgo any of her future options by accepting the offer. In short, if she accepts, then Alice loses no opportunities she would have had otherwise, nor does she make any of the outcomes of any later choices worse under the different states of the world she's uncertain about. In this kind of case, we can say that her learning is genuinely *cost-free*. And on the basis of several quite general formal results, we have known for a long time that it is always rational for a good Bayesian agent to pursue genuinely cost-free learning.[13]

But truly cost-free learning is rare indeed. Outside of purely fictional cases and artificial experimental situations, learning and inquiry usually involve some cost in effort, resources, time, and/or future opportunities. And often those costs can be considerable (as anyone paying back student loans will appreciate). So let us therefore consider a case of costly learning, which (we believe) is more closely analogous to the case of moral nihilism.

> **Case 2: The Movie.** Alice loves movies which have a "big twist," but only if she doesn't see the twist coming—if she were to know what twist is coming, then watching the movie would be worse than watching nothing at all. Luckily, when she is watching a movie with a twist, she only sees the twist coming about 20 percent of the time. (She vigorously avoids watching movies she's seen before if they have a twist.) Of those movies which don't have a twist, she usually considers them just so-so: better than nothing, but also not great.
>
> Alice is trying now to decide whether to watch a new movie; the alternative is to watch nothing. She knows nothing about the new movie, and she's 50/50 on whether it will have a twist. As she is making up her mind, the oracle appears again and offers to tell her the entire plot, free of charge. Should Alice accept?

Again, it is obvious what Alice should do. If Alice accepts the oracle's offer, then from her epistemic perspective there's a 50 percent probability that she'll learn the plot of a movie which has a twist, and hence she'll see

13. See, e.g., Irving John Good, "On the Principle of Total Evidence," *British Journal for the Philosophy of Science* 17 (1967): 319–21. Where Alice's degrees of belief are imprecise, matters are slightly more complicated, though the basic point still holds; for discussion, see Seamus Bradley and Katie Steele, "Can Free Evidence Be Bad? Value of Information for the Imprecise Probabilist," *Philosophy of Science* 83 (2016): 1–28.

the twist coming regardless of whether it was antecedently predictable or not. If that happens, she knows she'd dislike the movie intensely. Moreover, she will have ruled out obtaining the best possible outcome: watching a movie with a twist she doesn't see coming. She has nothing to gain and everything to lose. The fact that the oracle's offer is free of charge doesn't mean that it's cost-free. The situation is represented in figure 2, where the expected utility of *Reject* (12) is greater than the expected utility of *Accept* (5). (The specific values we've chosen for the utilities make no difference to the result, which in this case depends only on the preference ordering over the outcomes.)

The reasoning we've been attributing to Alice in these cases usually goes by the name *backward induction*. Effectively, while deciding whether to accept the oracle's offers of more information, Alice is treating her future self as a second player in a two-player game, whose later choice

a) determines Alice's outcomes now and
b) depends partly on (i) Alice's own present choice (*Accept* or *Reject*) and potentially also (ii) an unknown state of the world (Twist or No Twist).

The backward-inductive reasoning has been lightly idealized: we've been implicitly assuming that Alice knows (i.e., with certainty) exactly what her future self's options, beliefs, and utilities will be, and exactly how she will choose under the circumstances she might end up in. With some added sophistication we could remove these idealizations, but in practice they typically don't make much difference for simple cases like those we're

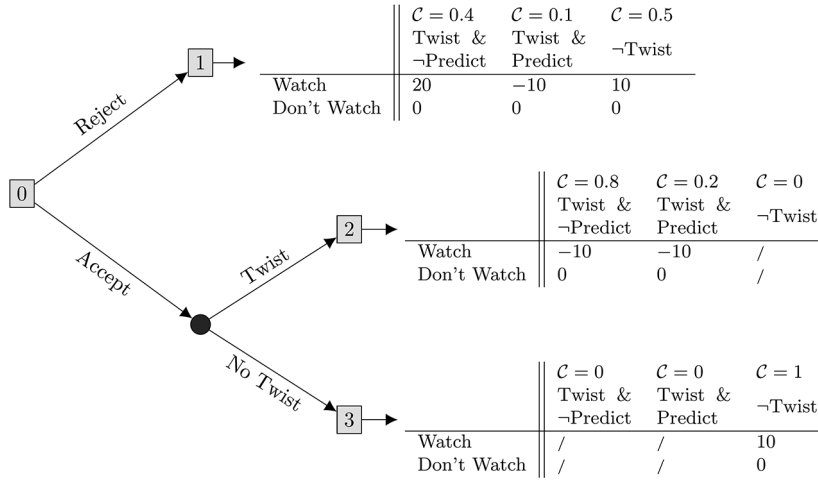

FIG. 2.—The Movie.

considering here. Insofar as Alice has a high degree of confidence on these matters, the main conclusions will remain more or less unchanged.

IV. THE VALUE OF MORALITY

Let us now apply some backward-inductive reasoning to learning about moral nihilism. We will begin our discussion with a toy case designed to bring out the basic structure of our argument and the core assumptions it rests on.

> **Case 3: The Sofa.** 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?

Before we say anything else, it must be emphasized that we're using 'warm fuzzy feeling' (henceforth *wff*) as a kind of placeholder. Our arguments do not rest on the idea that agents like Alice literally experience any sort of pleasurable sensation or violent passion when they act on their moral convictions. Our use of '*wff*' may equally well denote (say) a sense of personal accomplishment, or meaningful achievement, or a disjunction of the above. You could usefully treat '*wff*' as a stipulative name, designating something Alice values which is specifically tied to her believing that her moral preferences have been satisfied. We will argue that there are indeed such things in Section VI.

Given that, the first key point to note is that the outcomes of Alice's choices will depend not only on what state of the world is actual but also on Alice's beliefs about which kind of world she is in. In our description of the case, we have said that Alice has a preference for doing the right thing (whatever that may be), but she also cares about several other things

besides—for example, she has a desire to help her friend and a fear of social reprobation, and at least a slight preference for the little *wff*. Let's refer to the former as Alice's *moral preference*—that is, her preference for doing the right thing as such, for being a good person, for improving the overall moral goodness of the world. And let's refer to the latter (disjunctive) kind as her *nonmoral preferences*. Now, whether moral nihilism is true makes a difference to whether her moral preferences are satisfied, but what she believes about the status of moral nihilism makes a difference to her *wff* (which is one factor in her nonmoral preferences). The remaining nonmoral factors depend primarily on whether she chooses to help, not on the truth of moral nihilism or Alice's beliefs regarding it.

So let's turn that observation into an argument that Alice rationally ought to reject the oracle's offer. First, we will simplify matters by grouping together the disjunction of realist views under the heading 'Realism', just as we've done for 'Nihilism'. Alice's degrees of belief for Realism (Nihilism) will thus be the sum of her degrees of belief for the more specific (and mutually exclusive) varieties of realism (nihilism) that she's aware of, and her utilities for the outcomes of her choices given Realism (Nihilism) will be a weighted average of her utilities for those outcomes conditional on those specific varieties.[14] Furthermore, we'll begin by supposing that Alice doesn't yet consider how her choice vis-à-vis the oracle's offer might affect her outcomes in more distant future decision situations—she's focused for now just on how it will affect the immediate outcomes of her decision.

One more simplification, and then we will move on: say that Alice believes that p if and only if $C(p) \geq 0.9$, and then symbolize the outcomes as follows:

- x = *Help* at a world where Realism is true, and Alice believes Realism;
- y = *Help* at a world where Nihilism is true, and Alice believes Realism;
- z = *Help* at a world where Nihilism is true, and Alice believes Nihilism;
- q = *Don't Help* at a world where Realism is true, and Alice believes Realism;

14. It's worth noting again that the kinds of realism and nihilism that matter most here will be those that Alice herself will have in mind. Professional metaethicists will have varying degrees of confidence regarding each of the many varieties of realism and nihilism that they can distinguish between—but we can't expect this kind of nuance for Alice, who probably isn't even going to be aware of most of these distinctions, and for whom relatively flat-footed varieties of realism and nihilism are much more likely to be at the forefront of her mind.

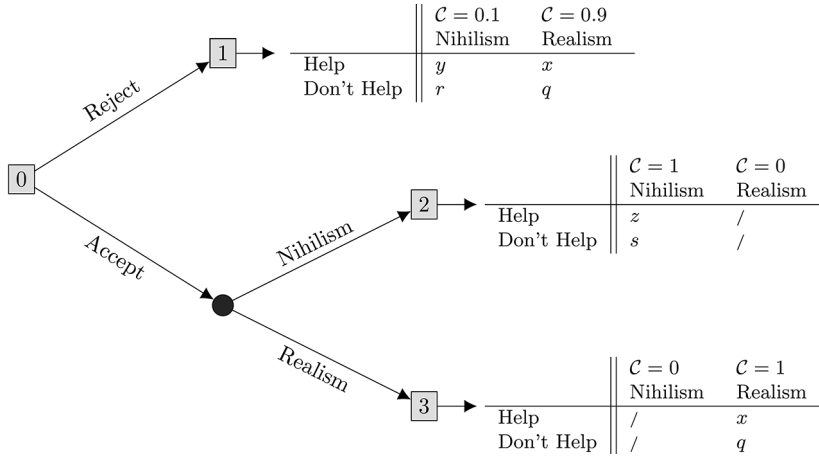

FIG. 3.—The Sofa.

- r = *Don't Help* at a world where Nihilism is true, and Alice believes Realism;
- s = *Don't Help* at a world where Nihilism is true, and Alice believes Nihilism.

The effect of all this is that helping/not helping at a world where Realism holds and $C(\text{Realism}) = 0.9$ has the same utility as helping/not helping at a world where Realism holds and $C(\text{Realism}) = 1$. This probably isn't exactly true, but it's likely to be approximately true, and it will let us present a relatively easy-to-follow version of our argument to begin with.

That's a fair few significant simplifications, and we promise we'll discuss desimplifying below, in Section VIII. Right now, though, we want to focus on the core of our argumentative strategy, and that core is best brought out by temporarily setting aside all the fiddly complications. Thus, we use figure 3 to represent Alice's decision situation in case 3.

Now, whether Alice ought to *Accept* or *Reject* the oracle's offer comes down to how we fill out the utility values at the different nodes of the decision tree. But suppose the following claims are true:

A1. At node 1, *Help* has maximal expected utility.
A2. x is at least as great as q.
A3. y is greater than either z or s.

Fill in the utility values any way you like consistent with A1–A3, and *Reject* will have strictly greater expected utility.[15]

15. *Proof*: where C_0 designates Alice's degrees of belief at node 0, A1 implies $EU_0(\textit{Reject}) = EU_1(\textit{Help}) = yC_0(\text{Nihilism}) + xC_0(\text{Realism})$, and from A2 and A3, $EU_0(\textit{Accept}) = [C_0(\text{Nihilism}) \times \max\{z, s\}] + xC_0(\text{Realism})$, where $y > \max\{z, s\}$; hence, $EU_0(\textit{Reject}) > EU_0(\textit{Accept})$.

Great—but why should we think that A1–A3 are true? Well, our core empirical assumptions are as follows:

> 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.

That is, prior to the oracle's offer Alice would prefer to do what she initially takes to be the right thing regardless of whether moral nihilism is true, because her moral and nonmoral preferences speak in favor of acting as such. Among the many factors that speak in favor of helping is the *wff*, though she would still choose to help even if she didn't expect to have it (Correlation). Furthermore, she thinks that she would no longer have the *wff* if she were to act the same way and did believe nihilism—if she no longer thinks there is a morally right thing to do, then she can't believe of herself that she has managed to do it. She'll still choose to act the same way, but she'd suffer the cost of doing so without having the *wff*, however small that cost may be (Cost). And, finally, that cost would not be compensated for by any gain to the utility of helping (or not helping) at a world where nihilism is true and she believes it. For example, if she cares about having true or more accurate beliefs, then she may get a bit of extra "epistemic utility" for having true beliefs about nihilism—but not enough to cover the costs of losing the *wff* (No Compensation). So Correlation directly supports assumptions A1 and A2, and given Correlation, Cost and No Compensation then jointly support A3.

Summarizing: our claim is that given her present belief state, Alice doesn't have much to gain by learning that nihilism is false—and, if it turns out that nihilism is true, then she has something to lose (the *wff*). Moreover, whatever she might gain by coming to have more true beliefs under that eventuality isn't enough to make up for what she stands to lose. Consequently, Alice should choose not to accept the oracle's offer of information. (We emphasize that our argument is concerned with this practical matter only, and not with the entirely separate epistemic question of what Alice ought to believe. Evidentialists have no cause for alarm!)

In the following sections, we'll provide empirical support for Correlation (Section V), Cost (Section VI), and No Compensation (Section VII).

V. CORRELATION

According to Correlation, people's moral and nonmoral concerns tend to correlate well with one another. We should emphasize from the outset that our arguments do not require that the correlation here be perfect. It should be all too obvious that it isn't; our moral and nonmoral concerns can and do sometimes come apart. Doing what is right may require sacrificing other things that we care about. Nonetheless, it is our contention that the correlation is fairly systematic and seems to hold true in general, even if it does not hold true on each and every occasion. We should also emphasize that we do not regard Correlation as a brute or necessary fact about moral agency. Our arguments only require that it amounts to a defensible generalization as regards typical human beings. The case that we sketch in what follows is therefore intended in an abductive spirit; given what we know about human evolution and psychology, it is reasonable to suppose that Correlation applies to a great deal many human beings. Clearly, more needs to be said to tie down precisely what we have in mind. We now get to work, beginning with the distinction between moral and nonmoral concerns.

As we will argue in greater detail below, people are generally invested in moral matters qua moral matters; they tend to care about acting in ways that they judge to be morally right, and about not acting in ways that they take to be morally egregious. But people are not only invested in moral matters. Most have other sorts of projects and interests as well. Some of these may be purely self-regarding (Alice may want to be popular). Others may be genuinely other-regarding (she may want to do something nice for her mother). Agents can no doubt have loftier concerns as well; Alice may enroll in university because she values knowledge for its own sake. (We attend to epistemic pursuits in greater detail in Sec. VII.) In practice, the boundaries between each of these can sometimes be fuzzy. However, we take it that nonmoral concerns can be meaningfully differentiated from concerns for moral rightness as such—ethics is not first philosophy.

Insofar as Alice's moral concerns correlate with nonmoral concerns such as these, the following will be true of her: whenever morality favors φ-ing, the balance of Alice's nonmoral preferences will independently tend to favor φ-ing as well, and whenever morality disfavors φ-ing, the balance of Alice's nonmoral preferences will independently tend to disfavor φ-ing as well, such that were morality removed from the equation, it would still be (ir)rational for Alice to φ. Correlation therefore rests on the well-worn idea that it's generally in one's interest to be (morally) good. We are

not inclined to regard this idea as especially controversial (the true battleground surely lies with the idea that it's necessarily in any agent's interest to be morally good). But we do not wish to take it for granted either. In what follows, we argue that morally good conduct generally promotes (what are reasonably regarded as) common human ends.

To begin with, humans care deeply about the welfare of others. Even young infants exhibit strong other-regarding concerns.[16] This is unsurprising. The survival of our species has long been predicated on successful cooperation, and there has long been biological and cultural selection for emotional responses that support it. We feel sympathy in response to others' suffering, anger in response to their selfishness, and guilt in response to our own. Each of these experiences has motivational import. Sympathy motivates helping behavior,[17] anger fuels punitive action,[18] and guilt encourages making amends.[19] It is telling that a callous disregard for others is among the key diagnostic criteria for a number of human pathologies—antisocial personality disorder and conduct disorder, for example.[20]

People also care about what others think of them. Social disapproval is often seen as an especially toxic form of punishment; many would prefer pain, jail time, amputation, or death to a heavily tarnished reputation.[21]

16. Ulf Liszkowski et al., "12- and 18-Month-Olds Point to Provide Information for Others," *Journal of Cognition and Development* 7 (2006): 173–87; Felix Warneken and Michael Tomasello, "Altruistic Helping in Human Infants and Young Chimpanzees," *Science* 311 (2006): 1301–3; Felix Warneken and Michael Tomasello, "Helping and Cooperation at 14 Months of Age," *Infancy* 11 (2007): 271–94; Robert Hepach, Amrisha Vaish, and Michael Tomasello, "A New Look at Children's Prosocial Motivation," *Infancy* 18 (2013): 67–90.

17. Daniel C. Batson et al., "Is Empathic Emotion a Source of Altruistic Motivation?," *Journal of Personality and Social Psychology* 40 (1981): 290–302; Daniel C. Batson et al., "Five Studies Testing Two New Egoistic Alternatives to the Empathy-Altruism Hypothesis," *Journal of Personality and Social Psychology* 55 (1988): 52–77.

18. Astrid Hopfensitz and Ernesto Reuben, "The Importance of Emotions for the Effectiveness of Social Punishment," *Economic Journal* 119 (2009): 1534–59; Rob Nelissen and Marcel Zeelenberg, "Moral Emotions as Determinants of Third-Party Punishment: Anger, Guilt and the Functions of Altruistic Sanctions," *Judgment and Decision Making* 4 (2009): 543–53; Elise C. Seip, Wilco W. Van Dijk, and Mark Rotteveel, "Anger Motivates Costly Punishment of Unfair Behavior," *Motivation and Emotion* 38 (2014): 578–88. It may be the case that anger has a different behavioral profile in other (non-Western) cultures; see Owen Flanagan, *The Geography of Morals: Varieties of Moral Possibility* (New York: Oxford University Press, 2016), 154.

19. Tamara J. Ferguson, Hedy Stegge, and Ilse Damhuis, "Children's Understanding of Guilt and Shame," *Child Development* 62 (1991): 827–39; June P. Tangney et al., "Communicative Functions of Shame and Guilt," in *Cooperation and Its Evolution*, ed. Kim Sterelny et al. (Cambridge, MA: MIT Press, 2013), 485–502.

20. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*-5) (Washington, DC: American Psychiatric Association, 2013).

21. Andrew J. Vonasch et al., "Death before Dishonor: Incurring Costs to Protect Moral Reputation," *Social Psychological and Personality Science* 9 (2018): 604–13.

Again, this is unsurprising for an ultrasocial and ultracooperative species. It is in our interests to secure the approval of others. It is arguably even more in our interests to avoid their disapproval. The price of being unpopular is high, ranging from lower job prospects to lower life expectancy.[22] Given the great costs of being disliked, some hypothesize that natural selection may have favored a robust concern for reputation.[23]

In light of evidence of this kind, we are inclined to regard Correlation as a reasonable empirical generalization. Morality centrally concerns the claims and interests of others, and we tend to care about how other people fare in life. Likewise, morally good conduct has clear reputational benefits. Most people favor generosity. Everybody appreciates compassion. Nobody likes an asshole.

Before moving on, we want to note that Correlation can be motivated on intuitive grounds as well. To this end, it helps to consider what it would take for it to be radically false. On the one hand, it may be that most people's nonmoral concerns are radically out of kilter with what morality requires of them. This would be true if, for example, most people had nothing but repugnance for their fellow travelers, deriving happiness from their pain, and caring little for their good opinion. In acting rightly, such agents would represent the (somewhat caricatured) Kantian ideal of moral agency: moral automatons propelled by the sheer force of the moral law. Alternatively, it may be that most individuals subscribe to excessively demanding moralities. The radical utilitarian may struggle to live up to her moral ideals, devoting most of her resources to her beloved children—all the while believing that morally she ought not to be doing so.

It would not necessarily be irrational for either of these agents to inquire after the truth of moral nihilism. Indeed, each stands to benefit should nihilism turn out to be true. Our Kantian would be free to act on her aversions and inclinations (however ignoble), the utilitarian no longer caught in the throes of excessive moral demands. We do not doubt that such agents exist.[24] But we do doubt their prevalence. Again, the empirical data here are telling; given what we know about human psychology, an individual completely devoid of fellow feeling is reasonably classified as anomalous. But the intuitive data are telling as well. Moral philosophers

22. Bruce Western, Jeffrey R. Kling, and David F. Weiman, "The Labor Market Consequences of Incarceration," *Crime and Delinquency* 47 (2001): 410–27; James S. House, Karl R. Landis, and Debra Umberson, "Social Relationships and Health," *Science* 241 (1988): 540–45.

23. See, e.g., Dan Sperber and Nicolas Baumard, "Moral Reputation: An Evolutionary and Cognitive Perspective," *Mind and Language* 27 (2012): 495–518.

24. Indeed, we were once asked (in a tone that suggested that the animating thought behind the question was all too obvious) why we did not seriously consider the best part of converting to moral nihilism—namely, the liberty to unabashedly pursue self-interest.

have long objected to Kantian and utilitarian ideals precisely because they are humanly unachievable.[25] People care deeply for those closest to them, and it is not reasonable to expect that such concerns can be surgically removed from their maxims or moral calculus. We agree. Any moral theory that proposes to divorce moral action from the affective network that underwrites it is unlikely to be user-friendly.

VI. COST

According to our second empirical assumption, there is a potential cost associated with inquiry into moral nihilism. The general idea here is that an ordinary human agent like Alice still gets something of value even in those worlds where moral nihilism is true, so long as she believes that she has done what is right—to wit, the *wff*.[26]

Buried in this general idea are two further assumptions concerning Alice's preferences:

> De Dicto Desires. Alice desires to do what is morally right, whatever that may be.
>
> De Dicto Dividends. There is an extra payoff specifically tied to Alice's believing herself to have satisfied her desire to do what is morally right.

We take the latter assumption to be the more controversial of the two, and so it is there that we will focus the majority of our critical attention. But let us offer some brief words of support for the first assumption.

De Dicto Desires attributes to Alice a standing desire to do the (morally) right thing, where that desire is given a *de dicto* reading: Alice desires to do the right thing, whatever that may be. Put differently, Alice desires to do the right thing as such; she wants to perform actions that are morally right under that description. She does not merely want to

25. Most famously, Michael Stocker, "The Schizophrenia of Modern Ethical Theories," *Journal of Philosophy* 73 (1977): 453–66; and Bernard Williams, "A Critique of Utilitarianism," in *Utilitarianism: For and Against*, ed. Bernard Williams and Jack J. C. Smart (Cambridge: Cambridge University Press, 1972), 77–150, 97–99.

26. It bears mentioning that the cost of believing moral nihilism may also depend on certain features of one's social environment, such as the distribution of moral nihilists in the wider population. Our arguments assume that Alice inhabits our world as it is—a world in which nihilists are a minority. But it is possible that our arguments could favor slightly different conclusions given the assumption that Alice inhabits a world in which moral nihilists constitute the majority—depending, of course, on how exactly one envisages such a world. While this is an interesting possibility, it is unfortunately not one that we can afford to explore in detail here.

do the right thing *de re*—to engage in behavior which, as it happens, is morally right. Some philosophers have questioned whether human agents generally are motivated in this way.[27] Others still have questioned whether they ought to be.[28] Regarding the latter complaint, it should be noted that much hostility to *de dicto* moral desires is really just hostility to the suggestion that they exhaust a moral agent's motivational resources.[29] It is therefore important to emphasize that we do not take Alice's *de dicto* moral desires to be the only force that motivates her, morally speaking. Indeed, Correlation predicts that Alice will have a number of *de re* moral desires as well—desires to help the global poor, promote peace in the Middle East, or save the whales, for example. Our arguments do not rest on any suspicious motivational monism.

We are now in a position to defend (what we take to be) the more interesting and controversial component of Cost. This was, recall, the assumption that there is a payoff specifically tied to Alice's belief that she has done the right thing. Some care is needed in spelling this out, for, on certain natural interpretations, it is neither interesting nor controversial.

On the one hand, it seems both empirically and phenomenologically obvious that people tend to feel good when they do good. The phenomenon of "warm glow" suggests that positive feelings often accompany prosocial behavior.[30] However, reverting to this truism won't suffice for our purposes. It is not sufficient that whenever Alice acts rightly, she experiences a *wff*. This *wff* needs to be tied to her belief that she has done what is morally right.[31] Otherwise, Cost won't be plausible—indeed, there won't be any cost to be paid at all. If Alice's *wff* is ultimately explained by her *de re* moral desires, then it is not something that she stands to lose upon coming to believe moral nihilism.

The all-important question therefore is, when feeling good accompanies doing good, is it because the agent believes that she has done the right thing as such? To our knowledge, there have been scarcely any

27. See, e.g., Nomy Arpaly, "Huckleberry Finn Revisited: Inverse Akrasia and Moral Ignorance," in *The Nature of Moral Responsibility: New Essays*, ed. Randolph Clarke, Michael Mckenna, and Angela Smith (New York: Oxford University Press, 2015), 141–56, 149.

28. See, e.g., Michael R. Smith, *The Moral Problem* (Oxford: Blackwell, 1994), 75–76.

29. David O. Brink, "Moral Motivation," *Ethics* 108 (1997): 4–32, 27–29.

30. See James Andreoni, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *Economic Journal* 100 (1990): 464–77; James Andreoni and John Miller, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica* 70 (2002): 737–53; Heidi Crumpler and Philip J. Grossman, "An Experimental Test of Warm Glow Giving," *Journal of Public Economics* 92 (2008): 1011–21.

31. We're here concerned with whether there's a *wff* tied to satisfying one's moral preferences. This is not to assume that there is no *wff* tied to satisfying one's nonmoral preferences. Indeed, it is precisely because the latter seems so plausible that the disentanglement problem arises.

philosophical expeditions into the empirical literature bearing on this question. But suggestive evidence is there. With a little effort and determination, the *wff* tied to (believing oneself to be) doing the right thing *de dicto* can be disentangled from any *wff* that may be tied to (believing oneself to be) doing the right thing *de re*.

We attend first to important work on moral identity and self-serving biases. It is a fact now widely recognized in psychology that people care deeply about their moral selves. Moral commitments are often described as "identity-defining"; they play an important role in defining who one is.[32] This idea has been borne out empirically in a variety of ways.[33] In what follows, we argue that there's good reason to take this moral self-conception to include desires with moral content—namely, desires to be a morally good person, or a person who does the morally right thing. Thus, an agent's moral self-conception is not just a matter of her aspiring to be a helpful person, or someone who promotes happiness. If we're correct, belief in moral nihilism has the potential to effect a radical upheaval in an agent's self-understanding, a loss to her sense of self.

The importance that people attach to their moral identity is reflected in the cognitive biases that support it. Self-serving biases are surprisingly common in the moral sphere. Almost everyone thinks that they are morally above average.[34] It may be tempting to chalk this up to a more

32. Darcia Narvaez and Daniel K. Lapsley, "Moral Identity, Moral Functioning, and the Development of Moral Character," *Psychology of Learning and Motivation* 50 (2009): 237–74, 243. The point here is not to deny that nonmoral qualities may also be essential to one's identity. Nor is it to deny that there is interpersonal variation in how central agents take their moral self to be—for discussion on this, see Karl Aquino and Americus Reed II, "The Self-Importance of Moral Identity," *Journal of Personality and Social Psychology* 83 (2002): 1423–40, 1423. However, one's self-conception is rarely limited to nonmoral attributes, and moral attributes are rarely inconsequential to who we take ourselves to be. Indeed, and as we will argue shortly, people often take the moral element of their identity to be especially important.

33. See, e.g., Augusto Blasi, "Moral Cognition and Moral Action: A Theoretical Perspective," *Developmental Review* 3 (1983): 178–210; Augusto Blasi, "Moral Identity: Its Role in Moral Functioning," in *Morality, Moral Behavior and Moral Development*, ed. William. M. Kurtines and Jacob L. Gewirtz (New York: Wiley, 1984), 128–39; Kristen R. Monroe and Connie Epperson, " 'But What Else Could I Do?' Choice, Identity and a Cognitive-Perceptual Theory of Ethical Political Behavior," *Political Psychology* 15 (1994): 201–26; Kristen R. Monroe, "Morality and a Sense of Self: The Importance of Identity and Categorization for Moral Action," *American Journal of Political Science* 45 (2001): 491–507; Kristen R. Monroe, "How Identity and Perspective Constrain Moral Choice," *International Political Science Review* 24 (2003): 405–25.

34. David M. Messick et al., "Why We Are Fairer Than Others," *Journal of Experimental Social Psychology* 21 (1985): 480–500; Wim Liebrand, David M. Messick, and Fred Wolters, "Why We Are Fairer Than Others: A Cross-Cultural Replication and Extension," *Journal of Experimental Social Psychology* 22 (1986): 590–604; George R. Goethals et al., "Fabricating and Ignoring Social Reality: Self-Serving Estimates of Consensus," *Relative Deprivation and*

general human tendency; people do, after all, tend to overestimate themselves. Yet this cannot be the whole story, for self-serving biases are selective. They are more pronounced in some domains than others. And they turn out to be especially pronounced in the moral domain. Whereas we strongly overestimate our moral credentials, we only weakly (if at all) overestimate our intelligence.

This asymmetry has been dubbed the *Muhammad Ali effect*, and the evidence seems to have converged on the following explanation for it.[35] On the one hand, people do well to have a high opinion of themselves. A little embellishment can be a good thing.[36] But there are limits. If one is to take a flattering self-portrait seriously, then that portrait must be credible.[37] Interestingly, it turns out that a misleading moral résumé tends to have more staying power than a misleading picture of one's cognitive potential. Even an agent who routinely reneges on her promises can hope to preserve a saintly image—by citing the greater good that was served by her actions, say. By way of contrast, it is difficult to persist in the illusion that one is a mathematical mastermind after having struggled to add up the dinner bill. The point is often framed in terms of verifiability: judgments of intelligence tend to be more publicly and objectively verifiable, whereas judgments of moral caliber are more subject to "interpretational or attributional ambiguity."[38] This explanation dovetails nicely with independent evidence suggesting that there is less room for self-serving maneuver where objectively verifiable abilities are concerned.[39]

Now for the philosophical takeaway. What we are concerned to emphasize is this: the explanation for the Muhammed Ali effect doesn't really have legs if the moral image that self-serving biases serve to protect

---

Social Comparison 4 (1986): 135–57; Scott T. Alison, David M. Messick, and George R. Goethals, "On Being Better but Not Smarter Than Others: The Muhammad Ali Effect," *Social Cognition* 7 (1989): 275–95; Paul 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.

35. See Alison, Messick, and Goethals, "On Being Better"; and Van Lange, "Being Better but Not Smarter." The effect's namesake defended his suspiciously poor performance on an army mental exam by remarking, "I only said I was the greatest, not the smartest." See Muhammad Ali, *The Greatest: My Own Story* (New York: Graymalkin Media, 2015), 129.

36. See Shelley E. Taylor and Jonathon D. Brown, "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin* 103 (1988): 193–210.

37. Daniel T. Gilbert and Joel Cooper, "Social Psychological Strategies of Self-Deception," in *Self-Deception and Self-Understanding*, ed. M. Martin (Lawrence: University Press of Kansas, 1985), 75–94.

38. Van Lange, "Being Better but Not Smarter," 692; see also Alison, Messick, and Goethals, "On Being Better."

39. See, e.g., Richard B. Felson, "Ambiguity and Bias in the Self-Concept," *Social Psychology Quarterly* 44 (1981): 64–69.

lacks a moral element—if it is merely an image of an individual who tends to help, for instance. Helpful behaviors are, after all, just as publicly and objectively verifiable as intelligent behaviors; it is hard to persist in the illusion that one is a helpful person if one never rises to the occasion when the opportunity presents itself. But the foregoing explanation does have legs if we suppose that the moral self-image is an image of an agent who does what is morally right. Given this supposition, the interpretational ambiguity of self-directed moral judgments makes sense; an agent who never seizes the opportunity to help may very well persist in believing that she is morally good—perhaps there is simply more important moral work to be done than attending to those in one's immediate vicinity. In prizing their moral identity, then, human agents plausibly prize being morally good people. But if moral nihilism is true, then this self-image quickly breaks down; there is no moral goodness for anyone to instantiate—no pride to be taken in one's moral accomplishments, nor any virtue to be cultivated throughout one's life. A belief in moral nihilism therefore poses a risk to an agent's sense of self.

We will now suggest that the costs of believing moral nihilism are greater still. Drawing on work in media psychology, we sketch a provisional case for thinking that there is a positive experiential aspect associated with believing one's moral preferences to be satisfied.

Moral judgment is heavily implicated in the consumption of dramatic entertainment. Moral assessment determines the extent to which a character is liked or disliked, and viewers also find a drama more enjoyable when characters get their just deserts.[40] Importantly, the latter phenomenon really does appear to be mediated by moral judgment. The evidence for this is fairly straightforward: vary the moral standards, and the enjoyment of dramatic entertainment will vary as well.

Studies involving children are particularly telling. Zillman and Bryant exposed four- and eight-year-olds to three fairy tales, which differed only in their portrayal of the villain's fate: (i) pardoned, (ii) proportionately punished, and (iii) excessively punished. These two populations were chosen for a reason. Children around four years of age have a penchant for

40. Dolf Zillmann and Jennings Bryant, "Viewer's Moral Sanction of Retribution in the Appreciation of Dramatic Presentations," *Journal of Experimental Social Psychology* 11 (1975): 572–82; Dolf Zillmann and Joanne Cantor, "A Disposition Theory of Humor and Mirth," in *Humor and Laughter: Theory, Research, and Application*, ed. Anthony J. Chapman and Hugh C. Foot (London: Wiley, 1976), 93–115; 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.

excessive retribution (the more suffering, the better).[41] By eight, they typically develop a preference for proportionate retaliation, which informs their sense of justice. In keeping with their hypothesis, Zillman and Bryant found that four-year-olds' enjoyment of the fairy tale increased with the severity of punishment, whereas eight-year-olds enjoyed the fairy tale most in the second condition. These results suggest that enjoyment was tied to the satisfaction of subjects' moral preferences. Similar studies conducted on children support this interpretation, as do studies involving adults.[42]

It's been a long journey. Let us summarize the foregoing defense of Cost and conclude with some words of caution. We have argued that (i) human agents have moral preferences and (ii) there are payoffs tied to the satisfaction of these preferences—payoffs that would no longer be available to them in a world where they believe the truth of moral nihilism. If this is right, then nihilistic belief comes at a price. It is a difficult question just how high that price is. This is likely to depend on (among other things) the extent to which the phenomena that we explore are representative of agents' moral experiences more generally. If, for example, there is a strong positive experiential aspect associated with the satisfaction of all sorts of moral preferences (i.e., not only preferences involving desert), then our case starts to look even stronger—likewise if cultivating a moral identity is the strongest prospect for injecting meaning into one's life.

We see no clear path to adjudicating the latter issue at present. Psychologists haven't tested the potential consequences of belief in moral nihilism directly—at least not to our knowledge. To some extent, then, the evidence for our empirical assumptions must be mined rather than handpicked. There is certainly room to debate the degree to which these results can be generalized and how high the relevant costs would be. Nonetheless—and we emphasize—there would be costs. Human beings do not merely attach credence to the idea that they inhabit a moral world. They also attach (at least some) utility to being in one.

Before moving on to a defense of No Compensation, it's worth addressing an important concern with what we've been saying. Some may worry that we've oversimplified things by assuming that recognizing the truth of moral nihilism merely involves believing it. This overlooks the sophisticated cognitive strategies that often accompany defenses of moral nihilism. Some nihilists adopt a conservationist strategy, which recommends outright belief in nihilism only in particular contexts (e.g., a philosophy classroom); in everyday life, nihilists are encouraged to hold onto their

41. Zillman and Bryant, "Viewer's Moral Sanction."

42. For examples, see Zillman and Cantor, "Disposition Theory"; Raney, "Punishing Media Criminals."

ordinary moral beliefs.[43] Others advocate a fictionalist approach, which advises nihilists to make-believe first-order moral propositions.[44]

It may be supposed that insofar as Alice adopts either of these strategies, there would be no threat to her *wff* and thus no cost to her coming to discover the truth of moral nihilism. If she is a conservationist, then she can continue to believe that she acts rightly in helping Bob and to experience a *wff* when she does so. Matters are more complicated as far as the fictionalist strategy is concerned, but there is evidence to suggest that moral make-belief is capable of engaging similar emotions to moral belief, including (perhaps) the *wff* to which our arguments appeal.[45]

Each of these is a well-developed response to the "What next?" question for moral nihilists, and we cannot hope to attempt a thorough assessment of them here. Still, it's worth noting why we do not take these proposals to be devastating for Cost. We'll begin with the fictionalist option. It is not implausible that make-belief can produce similar emotional experiences to belief. (Few feel warm and fuzzy inside when watching man-eating spiders on their television.) But whatever affective responses make-beliefs are capable of eliciting, these seem importantly different from the affective responses triggered by belief.[46] One who believes that there is a poisonous spider lurking somewhere in their bedroom is apt to feel a very real kind of fear—a fear that someone who make-believes that a rock is a spider is unlikely to experience. Thus, it seems that the fictionalist option would at best soften the blow for Alice, in virtue of preserving something like a *wff*—but there is still a cost.

Now to the conservationist. Even assuming that it is possible for Alice to hold onto beliefs that she knows to be false—a claim with which many would take issue—her belief that moral nihilism is true may very well cross her mind in everyday contexts. She may, for example, find herself thinking that helping Bob with his sofa would not really be the right thing to do at all (since nothing is right). This is, after all, something that she believes, and it seems relevant to deciding whether or not she ought to put a strain on her back. Attending to this belief would, however, likely prevent Alice from experiencing the *wff* that she usually experiences when she does what she believes to be right. Importantly, our suggestion is not that moral nihilism would always be on Alice's mind. The point is simply that there's no reason to think that it wouldn't cross her mind on at least some occasions (outside the philosophy classroom). Even if Alice (qua conservationist) does sometimes experience a *wff*, these experiences are

43. See, e.g., Olson, *Moral Error Theory*.

44. See, e.g., Joyce, *Myth of Morality*.

45. Ibid., 197.

46. See Shaun Nichols, "Just the Imagination: Why Imagining Doesn't Behave Like Believing," *Mind and Language* 21 (2006): 459–74; Shaun Nichols and Stephen Stich, "A Cognitive Theory of Pretense," *Cognition* 74 (2000): 115–47.

likely to be far less reliable than those of her realist analogue. For these reasons, we are inclined to think that adopting a post-nihilist strategy would at best reduce the costs of nihilist belief for Alice; it would not remove those costs completely.

VII. NO COMPENSATION

We have argued that there is something that Alice stands to lose should she inquire after the truth of moral nihilism. This does not itself establish that it would be irrational for Alice to accept the oracle's offer, for there may also be something that she stands to gain. Nihilistic belief may come at a price, but perhaps that price is worth paying. It is our contention that this is unlikely to be true for an ordinary human agent like Alice. Whatever compensation (if any) she receives, it won't suffice to offset the costs of nihilistic belief. We'll now consider some challenges to this claim: objections from Goody Two-shoes, philosophers, and the value of true belief.

(We should like to remind the reader that we have already dispatched one potential line of resistance: the idea that Alice would be liberated from the shackles of morality. Insofar as Correlation is true, morally recommended actions are, for the most part, actions that Alice has good independent reason to pursue. Thus, it's not as though, upon coming to believe moral nihilism, Alice will finally be free to do what she really wants.)

The first challenge to No Compensation comes in the form of an overzealous Goody Two-shoes. It remains possible that Alice is excessively scrupulous. Perhaps she is extremely guilt prone, carrying the weight of the world on her shoulders following even the slightest misdemeanor. Or maybe she is extremely anxious about doing what is right, planning her schedule well in advance to minimize the potential for any moral mishaps. To be sure, Correlation may still apply to Alice—her nonmoral preferences may tend to favor acting in ways that are morally desirable. But if moral nihilism is true, then she can continue to do so without the associated guilt and anxiety. Presumably, these unpleasant experiences are something she can do without. And if they're unpleasant enough, then it may be worth her while to forgo any *wff* in order that she may finally be rid of them.

We doubt that the moral inner life of Alice's goody-goody counterpart is representative of human agents more generally. Guilt and anxiety may very well be features of our moral experience. But they are unlikely to be anywhere near as pervasive for an ordinary human being as they are for a relentless Goody Two-shoes. Here, it is instructive to consider real people who do satisfy the above description: those who suffer from *scrupulosity*. Scrupulosity patients exaggerate the moral gravity of their behavior, are often paralyzed with moral indecision, and regularly revisit

and scrutinize their moral past.[47] For a very select group of individuals, then, the goody-two-shoes challenge applies; scrupulosity patients would be well-advised to inquire after the truth of moral nihilism. But insofar as such persons form a small pathological population, we are not inclined to regard them as a threat to our generalization.

There is, of course, room for disagreement on this score. Some—most notably, moral abolitionists—associate moral belief with a range of psychological afflictions: from "guilt complexes" to "ego competition."[48] While we do not deny that moral life has associated emotional costs, we do not see any reason to take them to be nearly as pervasive as Hinckfuss does. Admittedly, it is difficult to fully decide this matter in the absence of more direct empirical evidence; to our knowledge, there are no studies comparing the emotional profiles of nihilists and moral believers. That said, we do take the classification of scrupulosity as a pathology as reason to regard more intrusive forms of moral guilt and anxiety as somewhat atypical.[49]

On a different note, it has been put to us that we ourselves are walking counterexamples to No Compensation. As we hinted earlier, philosophers are likely to have strong incentives to inquire after the truth of moral nihilism. They may want to attend an upcoming conference, or publish a paper on the topic. Or (being philosophers) they may simply enjoy pondering life's great questions. For a philosopher, even a lifetime's worth of *wff*'s may be a small price to pay for news on the nihilist front. We are open to this possibility. However, it must be admitted that philosophers aren't representative of the general human population. Indeed, they are grossly unrepresentative. So it is no threat to our arguments if philosophers ought to learn more about moral nihilism. Insofar as there is an exception here, it is surely a principled one.

A final challenge alleges that Alice may simply value having true beliefs for its own sake.[50] Humans are curious creatures. Sometimes, we just

47. For an excellent overview, see Chris H. Miller and Dawson W. Hedges, "Scrupulosity Disorder: An Overview and Introductory Analysis," *Journal of Anxiety Disorders* 22 (2008): 1042–58. For a fictional example, see Chidi from the television show *The Good Place*.

48. Ian Hinckfuss, *The Moral Society: Its Structure and Effects* (Canberra: Department of Philosophy, Australian National University, 1987), v. Many abolitionists are, of course, concerned about the dangers of moral practice to society more generally, not only the disadvantages that accrue to individuals. We restrict our focus to the latter concern here since it is more pertinent to our arguments.

49. For responses to Hinckfuss's arguments, see Joyce, *Myth of Morality*, chap. 7; Stephen Ingram, "After Moral Error Theory, after Moral Realism," *Southern Journal of Philosophy* 53 (2015): 227–48; Jessica Isserow, "Minimizing the Misuse of Morality," in *The End of Morality: Taking Moral Abolitionism Seriously*, ed. Richard Joyce and Richard Garner (New York: Routledge, 2018), 131–49.

50. This is importantly distinct from the claim that truth has value for its own sake. What's important given how we're understanding utility is how truth figures in an agent's preferences, whether the agent herself cares about having true or more accurate beliefs.

want to know the truth about things, independently of any ends this may serve.[51] Are there more than ten thousand light bulbs in the Sydney Opera House? How many blades of grass were in the Hanging Gardens of Babylon? What is the exact number of wildebeest currently sweeping majestically across the African plains? These seem like pointless questions; their answers would be of no obvious practical benefit to Alice. Nonetheless, she may still prefer to learn them. The same may be true of moral nihilism; whatever Alice loses in *wff*'s may be compensated for in the currency of true beliefs.

We are not persuaded. The transition from human curiosity to an intrinsic concern for truth ought to be viewed with a healthy suspicion. The answers to seemingly pointless questions may very well have non-obvious benefits for Alice. She may want closure (think about having to miss the end of an exciting soccer match). She may enjoy entertaining her friends with surprising tidbits over dinner. Alternatively, she may subscribe to a better-safe-than-sorry policy. (What if she one day had to bet on the number of light bulbs in the Sydney Opera House?) Following Michael Brady, an agent's interest in the truth may be exhausted by her interest in some "unacknowledged practical goal."[52] Curiosity need not reflect an interest in the truth as such.

Of course, these confounding factors can be stipulated away. Suppose now that Alice must choose between two worlds, ω_1 and ω_2, which differ only in the following respect: at ω_2, Alice believes some pointless truth *T*. Suppose further that *T* really is pointless: none of the outcomes of Alice's choices will ever hang upon it. For our part, we see nothing to clearly recommend one world over the other. It certainly wouldn't strike us as bizarre if someone were indifferent between the two. That being said, it wouldn't strike us as bizarre if an agent preferred ω_2 to ω_1 either. After all, preference is cheap; even very small factors can make a difference when nothing else is at stake. What we do want to emphasize, however, is this: inasmuch as there is a preference here, it is at best an extremely miniscule one. Insofar as ordinary human agents like Alice do value truth, it's not clear that they value it very much.[53] Alice may prefer to pursue the truth when literally nothing else is at stake. But if this is truly the best that truth can do, then No Compensation remains in good

51. See Alvin I. Goldman, *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986), 98; Jonathan L. Kvanvig, *The Value of Knowledge and the Pursuit of Understanding* (Cambridge: Cambridge University Press, 2003), 41.

52. Michael Brady, "Curiosity and the Value of Truth," in *Epistemic Value*, ed. Adrian Haddock, Alan Millar, and Duncan Prichard (Oxford: Oxford University Press, 2009), 265–83, 270.

53. Chase Wrenn uses these considerations to support a parallel claim about truth's intrinsic value; at best, truth is the least valuable intrinsic good. See Chase Wrenn, "Truth Is Not (Very) Intrinsically Valuable," *Pacific Philosophical Quarterly* 98 (2017): 108–28.

stead, for there clearly is something at stake when it comes to belief in moral nihilism.

VIII. OBJECTIONS AND DESIMPLIFICATIONS

We promised we would say a bit more about some of the simplifications made in the argument of Section IV, and there are also a small number of objections we'd like to discuss, so we'll do that before concluding.

A. On Rationally Doubting Moral Nihilism

To start with, one might worry that given the possibility of metaphysically or epistemically necessary nihilist theories, our background assumption that it's both possible and rationally permissible that $0 < C(\text{Nihilism}) < 1$ might run up against our other Bayesian assumptions about Alice's beliefs.

Specifically: it is usual to define C over an algebra of propositions drawn from an underlying space of worlds Ω, where inconsistent propositions are modeled as the empty set of worlds. Given this, and since we've explicitly assumed that $C(p) = 0$ and $C(\neg p) = 1$ if p is inconsistent, C must assign 1 to anything that is true at all worlds in Ω. So if either

a) Ω is the space of metaphysically possible worlds and there is at least one variety of moral nihilism which happens to be true as a matter of metaphysical necessity or

b) Ω is the space of "a priori possible" worlds and there is at least one variety of moral nihilism which happens to be true a priori,

then the disjunction of moral nihilist theories will be true at all worlds in Ω, and $C(\text{Nihilism}) = 1$.

We think such worries would be misplaced. In particular, if a particular choice of Ω makes rational doubt regarding moral nihilism impossible or irrational, then we have a good reason to choose a different Ω. A nonideal agent can have very good reasons for being less than certain of a metaphysical necessity, or even an a priori truth if it's sufficiently nonobvious, without thereby being labeled irrational by any standard of rationality that's reasonably applicable to ordinary human beings. In connection with this, we note that the truth or falsity of (the disjunction of specific varieties of) moral nihilism is not entailed by classical logic—so there would be no problem of the kind being discussed here if we just let Ω be the space of classically logically possible worlds, and we see no reason not to think of Ω in this way.[54]

54. One of the authors would also like to add that this problem will arise only if a metaphysically or epistemically necessary moral nihilism is true—so isn't it lucky, then, that the most plausible versions of moral nihilism are both metaphysically and epistemically contingent!

B. On Rationally Believing Moral Nihilism

Our investigation is concerned with the rational advisability of learning—and thus, presumably, coming to believe—the truth of moral nihilism. Yet some may question whether genuine belief in moral nihilism is possible—that is, whether even nihilists themselves can truly be said to Believe it with a capital *B*. There is, after all, more to believing a theory than merely asserting it, being reasonably confident in it, or defending it in print. Arguably, someone who truly Believed in moral nihilism would neither enlist moral concepts in her deliberations nor experience familiar moral emotions. Yet this may seem like a tall order for even the most hard going of moral nihilists. Some readers may therefore be sympathetic to Charles Pigden's suggestion that those of a nihilistic persuasion can at best aspire to an "inconsistent moral nihilism," espousing but not truly Believing their own theories.[55] But this would seem to spell trouble for our arguments, since they may seem to require that Alice would Believe moral nihilism upon coming to learn of its truth.

However, and despite what initial appearances may suggest, nothing that we say depends on the assumption that Alice really would Believe moral nihilism, if the oracle were to tell her that it is true. According to the decision-theoretic framework that we're assuming, the rationality of a particular choice depends on an agent's own conception of the likely consequences of that choice—not on the actual consequences of that choice. Thus, the question is not, "Would Alice Believe?" It is rather, "Does Alice believe that she would Believe?" And insofar as Alice does view Belief in this more full-blooded sense as the likely consequence of being told that moral nihilism is true by a source she knows is entirely reliable, then that is enough for our arguments to proceed.

Still, some may want to press: is there any reason to think that Alice could genuinely Believe moral nihilism? There is in fact suggestive empirical evidence which speaks to this possibility. While no studies have, to our knowledge, been conducted on moral nihilists themselves, some lessons can be drawn from empirical investigations into atheism. Like moral belief, religious belief seems to be deeply ingrained: it is inculcated early in development, is socially supported, and builds on phylogenetically ancient features of our cognitive architecture.[56] Nonetheless, the evidence

55. Charles Pigden, *Non-naturalism versus Nihilism: Coursebook* (Otago: University of Otago, 1991); cited in Daniel Nolan, Greg Restall, and Caroline West, "Moral Fictionalism versus the Rest," *Australasian Journal of Philosophy* 83 (2005): 307–30.

56. See Pascal Boyer, *And Man Creates God: Religion Explained* (New York: Basic, 2001); Justin L. Barrett, "Exploring the Natural Foundations of Religion," *Trends in Cognitive Sciences* 4 (2000): 29–34; Scott Atran and Ara Norenzayan, "Religion's Evolutionary Landscape: Counterintuition, Commitment, Compassion, Communion," *Behavioral and Brain Sciences* 27 (2004): 713–30.

suggests that the epistemic transition from theism to atheism is possible.[57] Importantly for our purposes, converted atheists don't merely cease asserting that a deity exists. Rather, there is a wider sea change in how they navigate their way around the world—they no longer engage in prayer, or attend church, for example—which suggests a bona fide change in Belief.[58]

Admittedly, religious belief and moral belief are not exactly analogous; any inferential traffic between the two must therefore proceed with caution. Still, and in the absence of direct empirical evidence bearing upon belief (or Belief) in moral nihilism itself, the case study of atheism provides defeasible reason to suppose that genuine nihilistic Belief is possible.

C. On Immediate versus Long-Term Effects

We have been supposing throughout that Alice considers only the immediate effects that accepting or rejecting the oracle's offer will have with respect to her decision whether to help Bob. This greatly oversimplifies her real decision situation: any minimally rational agent like Alice shouldn't be considering only how a change in her information state might affect her outcomes at 10:00 a.m. on Tuesday, March 15, when Bob is attempting to enlist her help. If she learns that moral nihilism is true now, then that change of belief is liable to have far-reaching implications for how well her preferences are satisfied in many of her future "moral" choices, and Alice ought to take these implications under consideration to the extent she is able.

Now, we cannot feasibly recreate in a simple decision tree all of the temporally downstream factors that might matter to Alice's decision that she is aware of. But we see Alice's decision whether to help Bob as representative of the kinds of moral choice situations that an agent is likely to face over the course of a lifetime. If so, then Alice needs to weigh up the average cost of losing the *wff* over the course of a lifetime, versus whatever benefits she may receive now or in the long run from learning more about the truth of nihilism. We therefore take it that our arguments in support of Correlation support the claim that the case of the sofa is representative, and that our arguments in support of Cost and No Compensation can be naturally adduced in favor of the claim that the (potential) loss of a *wff* over the course of a lifetime isn't worth whatever benefits come with having more true beliefs about nihilism.

57. For a review, see Ara Norenzayan and Will M. Gervais, "The Origins of Religious Disbelief," *Trends in Cognitive Sciences* 17 (2013): 20–25.

58. See Jesse M. Smith, "Becoming an Atheist in America: Constructing Identity and Meaning from the Rejection of Theism," *Sociology of Religion* 72 (2011): 215–37.

D. On Belief versus Certainty

The other main complicating factor is not so easy to deal with. In Section IV we assumed that there's no difference in the utility that attaches to *Help/Don't Help* when $C(\text{Realism}) = 0.9$ and when $C(\text{Realism}) = 1$, at those worlds where Realism holds. The main upshot of this was that we could use the values x and q at both nodes 1 and 3 in figure 3. However, if this simplifying assumption is false, then figure 3 is a misrepresentation, and our earlier formal argument needs to be generalized to accommodate the possibility that the utilities of those outcomes might be different.

Hence, let x^* designate the value of *Help* at worlds where Realism is true and Alice is certain that it's true, and let q^* designate the value of *Don't Help* at the same kind of worlds. Given this, we'll get the same result in favor of Alice choosing *Reject* if we keep assumption A1 as is and replace assumptions A2 and A3 with[59]

A2* $x^* \geq q^*$.
A3* $[(x^* - x) \times C_0(\text{Realism})] < [(y - \max\{z, s\}) \times C_0(\text{Nihilism})]$.

A2* should be more or less just as plausible as A2. After all, if *Help* is more valuable than *Don't Help* at worlds where Realism is true and $C(\text{Realism}) = 0.9$, then it will likely also be more valuable when $C(\text{Realism}) = 1$. So we take it that Correlation is already enough to support A2*.

The harder one to justify is A3*. What it says is not at all easy to put into plain English. Nevertheless (take a deep breath): if there's any increase/decrease to the utility of *Help* at Realism-worlds that would result from a shift from $C(\text{Realism}) = 0.9$ to $C(\text{Realism}) = 1$ (weighted by Alice's original degree of belief in Realism), then that increase/decrease is less/greater than any decrease/increase in the utility of the optimal choice at Nihilism-worlds that would result from a shift from $C(\text{Nihilism}) = 0.1$ to $C(\text{Nihilism}) = 1$ (weighted by Alice's original degree of belief in Nihilism).

If we assume that $y > z > s$ and $x^* > x$, then there is a straightforward consequence of A3* for the purposes of our argument: the higher $C_0(\text{Realism})$ is, the less any cost (the difference between y and z) matters,

59. *Proof*: Let $C_0(\text{Nihilism}) = a$ and $C_0(\text{Realism}) = b$. From A1, $EU_0(\textit{Reject}) = ay + bx$, and from A2*, $EU_0(\textit{Accept}) = au + bx^*$, where u designates whichever is the largest of z or s. Now $bx^* - bx = b(x^* - x)$, which rearranged is $bx^* = bx + b(x^* - x)$. Likewise, $ay - au = a(y - u)$; hence, $au = ay - a(y - u)$. Thus, $EU_0(\textit{Accept}) = ay - a(y - u) + bx + b(x^* - x)$. Or in other words, $EU_0(\textit{Accept}) = EU_0(\textit{Reject}) - a(y - u) + b(x^* - x)$. By A3*, $a(y - u)$ is more than $b(x^* - x)$, so $EU_0(\textit{Accept}) < EU_0(\textit{Reject})$.

and the more any gain (the difference between x^* and x) matters. So, for example, to get the conclusion that Alice should choose *Reject*,

- If C_0(Realism) = 1/2, the cost must be more than the gain.
- If C_0(Realism) = 2/3, the cost must be more than 2 times the gain.
- If C_0(Realism) = 9/10, the cost must be more than 9 times the gain.
- If C_0(Realism) = 99/100, the cost must be more than 99 times the gain.

The upshot here is that there is a further dimension that needs to be considered before we can draw any conclusions about Alice's decision: if there's a difference between x^* and x, then Alice's initial degree of belief in Realism matters.

So what reason would we have for thinking that there's a difference between x^* and x? We have already argued that having slightly more accurate beliefs isn't (very) valuable for its own sake, so at most there's only a tiny gain in this case that might come from "epistemic utility." But perhaps Alice also gets more of a *wff* (or a more valuable *wff*) for doing what she thinks is the right thing, the more confident she is that there is a right thing to do. This seems plausible enough, so we're happy to take the suggestion on board. The question now is how this will impact our conclusion.

We're inclined to think that it won't make too much difference. In particular, it's likely that the higher C_0(Realism) is,

a) the smaller the difference between x^* and x, and
b) the greater the difference between y and z.

That is, if there is a positive correlation between the amount of *wff* and one's confidence in Realism, then we would expect very small increases in one's confidence to generate correspondingly small changes in the amount of *wff*. Moreover, we would expect that the correlation isn't perfectly linear—in particular, we think it's plausible that after a certain point, increasing one's confidence in Realism comes with diminishing returns in increases in the amount of *wff*. As one's confidence in Realism gets very high, there won't be much difference to the amount of *wff* that arises from becoming certain of Realism. And on the flip side, the higher C_0(Realism) is, the more *wff* there is to lose from becoming certain in Nihilism. Thus, if C_0(Realism) is quite high indeed—say, 99/100—then we would expect the difference between x^* and x to be correspondingly tiny and the difference between y and z to be much larger. And a cost that is 99 times larger than a very small amount need not be very large at all.

We have represented Alice's initial belief in Realism at 90 percent confidence. We don't have any strong empirical evidence to back that

up—it's a rough estimate supported by tutorial discussions with undergraduate philosophy students in ethics and metaethics classes in English-speaking universities. Make of that what you will. But the point just raised is quite general. While it's true that the higher C_0(Realism) is, the greater the difference between the cost and the gain must be for our argument to go through, there's also a trade-off given that the higher C_0(Realism) is, the greater the cost and the less the gain. In light of that trade-off, and in conjunction with the points raised in Sections V–VII, we take it that there's a strong case that Alice (probably) ought to reject.

IX. CONCLUSION

We have argued that nihilistic belief is not cost-free for an ordinary human agent like Alice. Moreover, this cost is unlikely to be offset by anything that an agent stands to gain in learning more about moral nihilism. True beliefs may count for something. But it is doubtful that any modicum of value they have would compensate for the costs of nihilistic belief over a lifetime. We have defended these assumptions about Alice's preferences on both intuitive and empirical grounds. If they are roughly correct, then the average person would be well-advised not to inquire after the truth of moral nihilism.

It is easy to mistake our conclusion for a Pascalian one. As one reader lamented, we seem to follow Pascal in recommending "pleasing lies over truth." However, nothing that we say supports this conclusion more generally. Presumably, there are many cases in which opting for convenient falsehoods would not maximize an agent's expected utility; in such cases, our reasoning would counsel pursuit of the truth. Some among our readership may have hoped for the result that it is never rational or advisable to pursue lies over truth. But if that's so, we suspect that their real beef is with the general theory of rationality that we assume, and a defense of that theory is not a burden that this article can reasonably be expected to bear.

Other readers may wonder whether our style of argument generalizes to the philosophical theories that they themselves hold dear. Given what we've said, is it similarly irrational to inquire after the truth of, say, modal realism or animalism? Well, it may be that similar arguments could be offered for these conclusions. We're not sure. What we do know is that such arguments would not be our argument, for our argument is built on evidence specific to moral nihilism. As far as we can tell, the empirical evidence that we've used to support (for example) Cost does not yield much if any insight into whether there is a payoff associated with believing that we are fundamentally persons who inhabit the only concrete world in the pluriverse. (Any animalists or modal realists who fear that a similar conclusion might apply to their own views can therefore

rest easy for now.) If the preceding discussion teaches us anything, it's that there are many details to consider. We therefore resist making any sweeping statements about whether premises analogous to Correlation, Cost, and No Compensation straightforwardly apply to other philosophical theories.

Though our conclusions apply in the first instance to ordinary people—a population from which philosophers have been swiftly and unceremoniously evicted—they do, of course, have implications for philosophers as well. According to a well-known tradition of thought, philosophers would sometimes do best to keep uncomfortable truths hidden from nonphilosophers or (as they are sometimes more affectionately known) "the folk."[60] And our arguments do suggest that moral nihilism is apt to be an uncomfortable truth. Some philosophers, then, may judge it best to keep their nihilistic opinions to themselves.

60. Sidgwick's so-called government house utilitarianism is a nice illustration. See Henry Sidgwick, *The Methods of Ethics* (London: Macmillan, 1984). For an application of this idea to moral nihilism, see Terence Cuneo and Sean Christy, "The Myth of Moral Fictionalism," in *New Waves in Metaethics*, ed. Michael Brady (Basingstoke: Palgrave MacMillan, 2011), 85–102. For further discussion of the utitilitarian case specifically, see Williams, "Critique of Utilitarianism"; and Katarzyna De Lazari-Radek and Peter Singer, "Secrecy in Consequentialism: A Defence of Esoteric Morality," *Ratio* 23 (2010): 34–58.