Journal of Ethics and Social Philosophy
VOL. 22, NO. 3 · SEPTEMBER 2022
https://doi.org/10.26556/jesp.v22i3.1266

EVOLUTION, UTILITARIANISM, AND NORMATIVE UNCERTAINTY

THE PRACTICAL SIGNIFICANCE OF DEBUNKING ARGUMENTS

Andreas L. Mogensen and William MacAskill

MANY PHILOSOPHERS believe that evolutionary considerations debunk whatever ethical beliefs they explain, drawing on the assumption that natural selection does not "track the truth" when it comes to ethics. If some evaluative disposition has been favored by selection—so the thought goes—then the truth value of any associated ethical belief is entirely irrelevant in explaining the fitness advantages associated with that disposition. Only by a coincidence could it turn out that these beliefs are true, and such a coincidence cannot reasonably be expected.[1]

Some philosophers who regard evolutionary explanations as debunking hold, in addition, that whereas evolutionary considerations provide discrediting explanations for the acceptance of many normative theories, they nonetheless cannot explain why utilitarians accept utilitarianism. Belief in utilitarianism seemingly transcends our evolved biases. Evolutionary considerations are thus thought to tip the balance in favor of utilitarianism by selectively debunking its competitors.[2]

The claim that natural selection cannot explain belief in utilitarianism is *prima facie* plausible. Utilitarianism asks us to attach equal value to the well-being of all individuals and act so as to maximally promote the general welfare. Given its complete impartiality and extreme demandingness, belief in utilitarianism would seem to represent a serious cost to an organism's inclusive fitness. Belief

1 See Joyce, *The Evolution of Morality*; Ruse, *Taking Darwin Seriously*; Street, "A Darwinian Dilemma for Realist Theories of Value." Strictly speaking, Street argues that natural selection explanations are debunking iff we assume meta-ethical realism.

2 Lazari-Radek and Singer, *The Point of View of the Universe*; Singer, *The Expanding Circle* and "Ethics and Intuitions"; Greene, "The Secret Joke of Kant's Soul"; Wiegman, "The Evolution of Retribution."

in utilitarianism may therefore be thought to have emerged *in spite of* the selection pressures shaping human moral psychology.

Our concern in this paper is with the possibility that evolutionary considerations still pose a serious problem for utilitarians. One particular concern, highlighted by Kahane, goes as follows.[3] Utilitarianism tells us to do whatever maximizes well-being. This prescription is empty unless we specify the nature of well-being. However, standard beliefs about well-being are prime candidates for evolutionary debunking. It is easy to see how natural selection would have led us to believe that pleasure is good for us and pain is bad for us. It is also easy to see how it could have led us to value desire satisfaction, or the characteristic ingredients in objective theories of well-being.[4] Since it looks like the beliefs we happen to hold about well-being will be debunked if any evaluative beliefs are, utilitarianism seems to be left without any practical content, even if the utilitarian principle is not itself undermined by evolutionary considerations.

We will argue that this is not the case. In sections 1 and 2, we show that successful debunking arguments targeting standard beliefs about well-being do not undermine the practical significance of utilitarianism, provided that we understand the requirements of practical rationality as sensitive to normative uncertainty.[5]

A different way in which evolutionary considerations may be thought to pose a serious problem for utilitarians is via the claim that belief in utilitarianism turns out to be debunked via the provision of a suitable evolutionary explanation for certain commonsense moral beliefs, since utilitarianism represents the reasoned extension of those beliefs, and so belief in utilitarianism is ultimately traceable to discredited starting points.[6] In section 3, we argue that evolutionary considerations may still increase the practical significance of utilitarianism even if belief in utilitarianism is debunked by evolutionary considerations, so long as belief in competing moral theories is undermined to an even greater extent.

3 Kahane, "Evolutionary Debunking Arguments" and "Evolution and Impartiality."

4 See Crisp, *Reasons and the Good*, 121–22.

5 Our response to Kahane therefore differs importantly from recent replies due to Bramble ("Evolutionary Debunking Arguments and Our Shared Hatred of Pain") and Jaquet ("Evolution and Utilitarianism"), who both try to resist the claim that relevant commonsense beliefs about well-being are debunked. We mean to show that the practical significance of utilitarianism is not undermined even granting that these beliefs are undermined. Obviously, this claim is compatible with the view that these beliefs are not in fact debunked.

6 Tersman, "The Reliability of Moral Intuitions"; Kahane, "Evolutionary Debunking Arguments."

1. DEBUNKING ARGUMENTS AND NORMATIVE UNCERTAINTY

To make our case, we will begin by clarifying how to conceptualize the damage done by evolutionary debunking arguments.

1.1. What Does It Mean for a Theory to Be Debunked?

Typically, the notion of debunking is characterized in terms of *categorical belief*: a theory is debunked iff belief in that theory is subject to an (undefeated) defeater.[7] But we could also characterize the notion of debunking in terms of *graded belief*.[8] We would then say that successful debunking arguments require us to (significantly) reduce our credence in various normative theories.

Plausibly, a debunking argument never requires us to reduce our confidence in some ethical theory to zero. To assign credence zero to some proposition is to be certain that one could never gain evidence that would raise one's credence above zero. But it would be extreme to suppose that debunking arguments could be so forceful as to render it impossible for any future evidence to support the normative theories we currently believe. Debunking arguments do not salt the earth.

Furthermore, we should not be certain of the soundness of any evolutionary debunking argument. Critics have alleged that these arguments rest on faulty epistemological principles, disputable meta-ethical presuppositions, and even mistakes about the nature of evolutionary explanations.[9] Thus, even if you are confident that some debunking argument is sound, you ought to assign non-negligible credence to the possibility that it is not.

1.2. Rational Decision-Making under Normative Uncertainty

It is plausible that we should never be completely certain of anything in ethics. Any reasonable person should acknowledge that their values could be mistaken and assign some degree of confidence to a range of ethical views. Since these different views will often diverge in what they tell us to do, we may wonder how we are to decide what to do, given our normative uncertainty.

7 Kahane, "Evolutionary Debunking Arguments"; Joyce, *The Evolution of Morality*.

8 As noted by Nichols, "Process Debunking and Ethics," 731.

9 For epistemological objections see White, "You Just Believe That Because …"; and Vavova, "Debunking Evolutionary Debunking." For meta-ethical objections, see Kahane, "Evolutionary Debunking Arguments." For objections from the philosophy of biology, see Mogensen, "Evolutionary Debunking Arguments and the Proximate/Ultimate Distinction" and "Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?"; and Hanson, "The Real Problem with Evolutionary Debunking Arguments."

One possible view is that we should be guided by the theory in which we are most confident.[10] In the literature, this view is known as *My Favorite Theory*. As it turns out, My Favorite Theory is beset with problems, the most troubling of which is that its recommendations are sensitive to arbitrary choices about theory individuation.[11] In recent years, a number of philosophers have argued that in cases of normative uncertainty we ought instead to act so as to *maximize expected choice-worthiness*.[12] This view is analogous with the orthodox decision-theoretic principle of maximizing expected utility.

Here is the basic idea. In a decision situation, an agent confronts a set of options. The agent's credence function assigns a probability to each member in a finite set of first-order normative theories, corresponding to the agent's confidence in the theory. A theory ranks the agent's options in terms of their choice-worthiness. We assume (for now) that choice-worthiness is interval-scale measurable and intertheoretically comparable. Roughly, this means that each theory tells us how much more (or less) choice-worthy one option is as compared to another and each theory can be represented as ranking the options according to the same scale of choice-worthiness. The expected choice-worthiness of some action is the sum of its choice-worthiness according to each of the theories in the set, weighted according to their probability. The most appropriate option is that which maximizes expected choice-worthiness.

Consider a stylized example. Suppose *S* is 70 percent confident that some form of rights-based deontology is true. According to this theory, it is wrong to intentionally harm one person in order to prevent two others from being harmed in the same way. *S* assigns the remainder of her confidence to utilitarianism.[13] An evil mastermind offers *S* the option to electrocute *A* in order to stop

10 Gracely, "On the Noncomparability of Judgments Made by Different Ethical Theories"; and Gustafsson and Torpman, "In Defence of My Favourite Theory."

11 See MacAskill and Ord, "Why Maximise Expected Choiceworthiness?" 332–35.

12 Lockhart, *Moral Uncertainty and Its Consequences*; MacAskill, *Normative Uncertainty*; Sepielli, "What to Do When You Don't Know What to Do." For objections, see Gustafsson and Torpman, "In Defence of My Favourite Theory"; Harman, "The Irrelevance of Moral Uncertainty"; and Weatherson, "Running Risks Morally." Our argument proceeds on the assumption that maximizing expected choice-worthiness accounts are at least approximately correct, at least in contexts where the different theories in which the decision maker is confident yield choice-worthiness values that are interval-scale measurable and intertheoretically comparable. For a recent, comprehensive defense of these assumptions, see MacAskill, Bykvist, and Ord, *Moral Uncertainty*.

13 There are obviously many different varieties of utilitarianism depending on what theory of welfare is adopted and how positive and negative welfare are weighted relative to one another. We note that since the choices in the example relate only to the minimization of suffering, classical and negative utilitarianism agree in their evaluation of this case. Throughout this

B and *C* from being electrocuted by the evil mastermind. Alternatively, she can refuse and allow *B* and *C* to be electrocuted. Her decision situation might then be represented as follows:

Matrix 1

	Deontology 70%	Utilitarianism 30%
Electrocute	5	25
Don't Electrocute	25	5

The numerical values in the cells represent the choice-worthiness scores of the different actions under the two moral theories. The deontological theory ranks Don't Electrocute as most choice-worthy. The utilitarian theory ranks Electrocute as equally choice-worthy. For simplicity, we assume that utilitarianism ranks Don't Electrocute as worse than Electrocute to the same extent that the deontological theory ranks Electrocute as worse than Don't Electrocute. Given these stipulations, the expected choice-worthiness of Electrocute is 11 and the expected choice-worthiness of Don't Electrocute is 19. Therefore, the most appropriate option in light of *S*'s confidence in the two moral theories is Don't Electrocute.

Decision matrix 1 assumed that electrocution harms a person, since it causes pain. *S* might not be totally certain that pain is intrinsically prudentially bad. To take account of this, we might think of *S* as distributing her credence over four different normative theories, each representing the conjunction of a moral theory and theory of well-being.[14] Assume that *S*'s confidence in utilitarianism remains at 30 percent and her confidence in deontology at 70 percent. Suppose, in addition, that she is 99 percent confident that pain is bad and 1 percent confident that pain is indifferent. Assuming for simplicity that the probability that pain is bad or indifferent is independent of which moral theory is true, the decision matrix might then look like this:

paper, we focus principally on cases like this, since the badness of suffering is focal in Kahane's discussion. *Mutatis mutandis*, our arguments can easily be transposed to deal with other putative sources of intrinsic subjective (dis)value, belief in which may be thought subject to evolutionary debunking arguments

14 Some readers may find it strange to think that utilitarianism can be combined with the view that pain is not bad, as utilitarianism may be understood by some to include certain beliefs about the nature of well-being, or to at least exclude views that treat pain as good or indifferent. Here we understand utilitarianism simply as the view that we ought to maximize aggregate well-being, and hence as compatible in principle with any theory of well-being.

Matrix 2

	Deontology Pain is bad 69.3%	*Utilitarianism* Pain is bad 29.7%	*Deontology* Pain is indifferent 0.7%	*Utilitarianism* Pain is indifferent 0.3%
Electrocute	5	25	15	15
Don't Electrocute	25	5	15	15

The right-hand side of decision matrix 2 looks as it does because we assume that if pain is neutral then either choice is equally permissible according to either theory. The side constraint against intentional harm has no force, since *A* is not harmed by electrocution. And there would be no reason to ensure that a smaller number of people are electrocuted on utilitarianism, since being electrocuted makes no difference to a person's well-being. Whatever *S* chooses will be equally unobjectionable, whichever moral theory happens to be true.

The prescription to maximize expected choice-worthiness still tells *S* not to electrocute. Its expected choice-worthiness is 18.96, compared to 11.04 for the alternative. Having some slight worry that pain is indifferent makes no difference to what is most appropriate for *S* to do in this context.

1.3. The Significance of Debunking Arguments

Suppose *S* becomes aware of a plausible evolutionary debunking argument that considerably reduces her confidence in deontology, but not in utilitarianism. Since utilitarianism has always seemed plausible to *S* apart from the fact that it conflicts with certain entrenched deontological intuitions, she becomes a lot more confident in utilitarianism. Suppose *S* now assigns 30 percent confidence to deontology and 70 percent confidence to utilitarianism. In that case, the expected choice-worthiness of Electrocute is 18.96, while the expected choice-worthiness of Don't Electrocute is 11.04. In that case, Electrocute is the most appropriate choice under normative uncertainty.

What if *S* is also made aware of a debunking argument targeting her belief that pain is bad? Well, if she loses all confidence in the badness of pain, this would mean that Electrocute and Don't Electrocute are equal in terms of expected choice-worthiness. In that case, the fact that she is also quite confident that utilitarianism is the correct moral theory would be genuinely irrelevant.

However, we have already ruled out the idea that debunking arguments require us to reduce our confidence to zero. Suppose, more realistically, that *S* ends up only 30 percent confident that pain is bad. In that case, the expected choice-worthiness of Electrocute is 16.2 and the expected choice-worthiness of Don't Electrocute is 13.8. Electrocute remains the most appropriate choice.

In fact, it should be straightforward to see that so long as *S* retains some con-

fidence in the badness of pain, reducing her confidence in this proposition to any arbitrary degree ultimately makes no difference to what would be most appropriate, given her relative confidence in utilitarianism vis-à-vis deontology. If pain is indifferent, then either action is equally choice-worthy no matter which moral theory is true. The normative theories represented in the right-hand side of the second decision matrix make no difference to the relative expected choice-worthiness of the two options. The question of which action is most choice-worthy in expectation is decided entirely by how *S* distributes her confidence across those normative theories on which pain is bad, represented in the left-hand side of the decision matrix. Therefore, so long as her relative confidence in utilitarianism is significantly greater, Electrocute remains the most appropriate option.[15]

Therefore, the availability of a debunking argument targeting the belief that pain is bad turns out to be without practical significance. As we recall, the debunking argument targeting *S*'s deontological moral intuitions did make a significant difference. In light of that argument, Electrocute became the most appropriate choice. And the fact that *S* is significantly more confident of utilitarianism ensures that this remains so regardless of the extent to which she reduces her confidence that pain is bad, so long as it remains above zero.

2. WHAT FOLLOWS?

Our discussion in the previous section focused on a stylized example, constructed using a number of simplifying assumptions. What does this case really tell us about our actual practical predicament?

2.1. Beyond Expected Choice-Worthiness

The example presumed that the normative theories to which *S* assigns credence yield choice-worthiness rankings that are interval-scale measurable and intertheoretically comparable. This might seem unrealistic.[16] Where these assumptions do not hold, we cannot act so as to maximize expected choice-worthiness. We have to apply some other rule.

Fortunately, this makes no difference to the key point for which we have argued. On any plausible principle for decision-making under normative uncertainty, the most appropriate option will be determined purely by *S*'s credence in those normative theories that assume the badness of pain. Her credence in those

15 Compare Ross, "Rejecting Ethical Deflationism," on the irrelevance of "uniform ethical theories" given normative uncertainty.

16 Gracely, "On the Noncomparability of Judgments Made by Different Ethical Theories"; and Ross, "Rejecting Ethical Deflationism."

theories that treat pain as indifferent will be irrelevant, since they treat her choice as indifferent. Only those theories that assume pain's badness can tip the balance.

By way of illustration, consider a principle that works for purely ordinal theories: the *Borda rule*.[17] According to the Borda rule, one option is more appropriate than another iff it receives a higher *credence-weighted Borda score*. An option's Borda score according to some theory is the number of options to which it is superior, minus the number of options to which it is inferior. Its credence-weighted Borda score is the sum of its Borda score under each theory multiplied by one's credence in the theory.

Suppose that deontology and utilitarianism did provide only an ordinal ranking of *S*'s options in terms of choice-worthiness. Given the previously stipulated confidence levels assigned by *S* to deontology, utilitarianism, pain's badness, and pain's indifference, her credence-weighted Borda-score for Electrocute is 0.12. For Don't Electrocute, it is −0.12. Electrocute is still most appropriate.

Furthermore, it is relatively easy to work out that the relative ranking of *S*'s options in terms of their credence-weighted Borda score is insensitive to her credence in pain's badness vis-à-vis its indifference, in that neither normative theory on which pain is indifferent contributes to the credence-weighted Borda score of either option. In this respect the Borda rule behaves just like the principle of maximizing expected choice-worthiness. And any other plausible principle should behave similarly.

2.2. Beyond Harm

Another respect in which the decision situation we have considered might be thought unrepresentative is that only the avoidance of harm was assumed to have normative significance.

However, a deontological theory might well posit that a rights violation occurs when one person electrocutes another without their consent, even if doing so is harmless. In that case, the deontological theory favors Don't Electrocute even on the assumption that pain is indifferent. *S*'s choice situation might then look like this:

Matrix 3

	Deontology Pain is bad 9%	*Utilitarianism* Pain is bad 21%	*Deontology* Pain is indifferent 21%	*Utilitarianism* Pain is indifferent 49%
Electrocute	5	25	10	15
Don't Electrocute	25	5	20	15

17 MacAskill, "Normative Uncertainty as a Voting Problem."

Here, the expected choice-worthiness of Electrocute remains highest. However, this can change if *S* becomes even more confident that pain is indifferent. Suppose she is 90 percent confident that pain is indifferent. Then the expected choice-worthiness of Electrocute becomes 14.05. The expected choice-worthiness of Don't Electrocute becomes 15.95. Don't Electrocute would then be most appropriate.

The reason for this should be clear. The utilitarian theory on which pain is indifferent does not tell for or against Electrocute. By contrast, the deontological theory on which pain is indifferent tells against. The more confident *S* becomes that pain is indifferent, the more weight she gives to these theories in deciding what to do. Since the utilitarian theory is indifferent on this point whereas the deontological theory is not, increasing her confidence that pain is indifferent strengthens her reasons for choosing Don't Electrocute.

It does not follow that the combined effect of a successful debunking argument targeting *S*'s deontological intuitions and another targeting her belief in the badness of pain will generally leave everything as it was before. This will hold true in some decision situations, but not in others. Whether things are left unchanged in any given case will be highly sensitive to the confidence *S* actually assigns to utilitarianism vis-à-vis deontology and to the badness of pain vis-à-vis its indifference. It will also be highly sensitive to the particular choice-worthiness ordering generated by each theory. This is easy to see by tinkering with the credences and rankings we used above. Slight adjustments can easily tip the balance.

It would be an astonishing coincidence if our credences and choice-worthiness rankings were calibrated so that reducing our confidence in deontology and in our beliefs about well-being never made any difference to which option was most appropriate in cases that potentially involve violation of side constraints. Furthermore, side constraints are just one point of contention between deontology and utilitarianism. Many of the remaining contrasts are purely a matter of how to weigh harms and benefits befalling different people. For example, deontological theories typically posit *agent-centered permissions*, in light of which each person is entitled to attach added weight to her own well-being. Deontological theories may also posit *irrelevant utilities*: a non-consequentialist might think it is more important to save a single individual from some terrible harm than provide a trivial benefit to each person in an arbitrarily large group of people.[18] The aggregative character of utilitarianism rules out this possibility.

In choice situations where agent-centered permissions or irrelevant utilities lead deontological theories to issue prescriptions that run against the implications of utilitarianism due to intertheoretic disagreement about the weighting

18 Kamm, *Morality, Mortality*, vol. 1.

of harms and benefits, reducing one's confidence in deontology will make an important practical difference, whereas reducing one's confidence that one's actions will make any difference to people's well-being will make no difference.

2.3. What about Really Bizarre Views?

A final worry centers on the possibility that debunking arguments require us to increase our credence in bizarre views about the nature of well-being. For example, we should perhaps increase our credence in the view that pain is intrinsically good for us and pleasure intrinsically bad, as we can be confident that this view would not have been selected for. But we have so far ignored this possibility.

In a similar vein, Kahane notes that certain highly counterintuitive beliefs about well-being will resist evolutionary explanation: "These would include the views that the good life consists of ascetic contemplation of deep philosophical truths, or celibate spiritual communion with God, or a kind of Nietzschean perfectionist aestheticism (which might even revel in pain), and so forth."[19] In combination with such theories, he notes, utilitarianism might retain its practical significance. However, its implications would be utterly repugnant: few people would be able to accept these implications. Is our argument vulnerable to this sort of worry? Does the ability of bizarre moral views to escape debunking mean that they are likely to end up playing a substantial role in determining what is most appropriate in light of our normative uncertainty?

That would be the case if evolutionary debunking arguments pushed our confidence in commonsense views about well-being down so far that it was not appreciably higher than our confidence in these wildly counterintuitive theories. We could end up in this position if debunking arguments required us to reduce our confidence in commonsense intuitions very close to zero. But the effect of encountering these arguments will not be so catastrophic. Debunking arguments may seem convincing, but it is far from certain that they are sound. For this reason, we ought to retain significant credence in commonsense views about well-being of which we were extremely confident prior to encountering these arguments. In the examples we considered earlier, we set *S*'s posterior credence in pain's badness at 30 percent or 10 percent. Given *S*'s antecedent confidence and the controversy surrounding the soundness of debunking arguments, even this might be too low.

If she is like the authors, *S* would have assigned a much, much lower prior probability to the view that pain is good or that celibate spiritual communion with God is the key determinant of well-being. Her posterior confidence in com-

19 Kahane, "Evolution and Impartiality," 334.

monsense views could therefore be orders of magnitude greater than her credence in wildly counterintuitive theories of this kind. The practical significance of these views would therefore be negligible.[20]

Of course, this would *not* be the case if her confidence in these counterintuitive theories should increase significantly upon encountering debunking arguments. That would be the case if one of these theories of well-being was like utilitarianism in that it seems plausible apart from the fact that it conflicts with certain entrenched commonsense intuitions that now get debunked, provided that the plausibility of the theory itself remains intact in the face of debunking arguments.

However, the theories considered here do not seem to fit that description. The view that pain is intrinsically good is not the sort of view that seems somewhat plausible, except for the fact that it conflicts with intuition. As we see it, it has basically zero inherent plausibility. The view that the good life is centered on celibacy, meditation, and prayer strikes us as false principally because it attaches value to things that seem valueless owing to our confidence that God does not exist. Debunking arguments will not change that fact.[21] We are more attracted to the view that contemplation of philosophical truths or the realization of aesthetic value can be intrinsic sources of well-being. Theories that count such goods as the primary or only determinants of well-being seem weird to us principally because they attach too little value to other things, such as pleasure or desire satisfaction. Nonetheless, these theories do not fit the criterion we specified above. To the extent that such theories have plausibility in light of the intuitive value of knowledge and aesthetic excellence, they will lose plausibility in the face of debunking arguments. After all, it is easy to see why natural selection should lead human beings to value knowledge: we are informavores by design.[22] There is also good reason to expect that natural selection has played a significant role in shaping our aesthetic responses.[23]

20 For the view that pain is good and pleasure is bad, there is a further argument for discounting its practical significance. When combined with utilitarianism, this view has exactly opposite recommendations to classical utilitarianism. Therefore, under normative uncertainty this theory simply "cancels out" part of one's credence in classical utilitarianism. For example, with 60 percent credence in deontology, 38 percent credence in classical utilitarianism, and 2 percent credence in pain-is-good utilitarianism, a rational decision maker will take the same actions as if she had 60 percent credence in deontology, 36 percent credence in classical utilitarianism, and 4 percent credence in a view that was indifferent between all options.

21 Except perhaps to increase our confidence in atheism; see Wilkins and Griffiths, "Evolutionary Debunking Arguments in Three Domains."

22 Dennett, *Consciousness Explained*, 176–82.

23 Dutton, *The Art Instinct.*

It might be argued that our confidence in the verdict that, say, pain is good should rise significantly once we are made aware of relevant evolutionary debunking arguments, simply because the belief that it is not the case that pain is intrinsically good has an evolutionary explanation. To the extent that evolutionary debunking arguments are sound, this ethical belief ought therefore to end up being debunked. In order to maintain probabilistic coherence, our credence that pain is good must rise accordingly, and so must rise significantly.[24]

We are not convinced by this line of argument. To see why, let us start by asking in what sense the belief that it is not the case that pain is intrinsically good can be said to have an evolutionary explanation. There are many things of which we are confident that they are not intrinsically good, such as having an odd number of hairs on one's left shin. In some sense, this confidence is explained in terms of evolution by natural selection, since "all phenotypes are to some extent the products of the process of evolution by natural selection."[25] Nonetheless, it is highly implausible to suppose that there existed some specific selection pressure that accounts for our confidence that it is not intrinsically good to have an odd number of hairs on one's left shin. Rather, we can presume that our confidence in this hypothesis is explained by considerations of parsimony, given the absence of any perceived reason to accept any contrary hypothesis. A similar story presumably accounts for our confidence that it is not the case that it is intrinsically bad to have an odd number of hairs on one's left shin. We are confident of these things on roughly the same grounds that we are confident that there is no luminiferous aether—because it is the simpler hypothesis.

On its face, beliefs such as that it is not intrinsically good (or bad) to have an odd number of hairs on one's left shin are not within the scope of evolutionary debunking arguments, precisely because they are not explained in terms of specific selection pressures yielding particular ethical intuitions and can instead be explained at a proximate level in terms of the application of a domain-general principle of parsimony.

We note, then, that we also find it implausible to suppose that there existed any specific selection pressure that accounts for our confidence that it is not the case that pain is intrinsically good—over and above whatever selection pressures account for the judgment that pain is intrinsically bad. With respect to the confidence previously assigned to that hypothesis, it is possible to redistribute that confidence over two alternative hypotheses: namely, that pain is intrinsically neutral and that pain is intrinsically good. The same principles of parsimony

24 We are grateful to an anonymous referee for this suggestion.

25 Brandon, *Adaptation and Environment*, 41.

that should lead us to be confident that it is neither intrinsically good nor intrinsically bad to have an odd number of hairs on one's left shin should presumably lead us to redistribute probability mass from the proposition that pain is intrinsically bad to the proposition that pain is intrinsically neutral, leaving our credence in the hypothesis that pain is intrinsically good effectively unchanged, remaining anchored at a very low prior.

3. UTILITARIANISM DEBUNKED?

Throughout sections 1 and 2, we have operated under the assumption that, whereas evolutionary considerations provide discrediting explanations for the acceptance of many normative theories, they nonetheless cannot explain why utilitarians accept utilitarianism. As a result, we have assumed that belief in utilitarianism is not debunked by evolutionary considerations. We have focused our attention on the worry that utilitarianism may nonetheless be robbed of its practical significance, given that our ordinary beliefs about the nature of well-being seem vulnerable to debunking arguments. In this final section, we briefly outline how our conclusions may nonetheless go through—and for roughly the same reasons—even if we grant that belief in utilitarianism can also be debunked.

We argued earlier that since belief in utilitarianism seems to represent a significant cost to an organism's inclusive fitness, belief in utilitarianism may be thought to have emerged in spite of—and not because of—the selection pressures shaping human moral psychology. A standard response to this suggestion is that we can explain belief in utilitarianism as the reasoned extension of the more restricted forms of benevolence and impartiality that we expect natural selection to have favored in the environment of evolutionary adaptedness, placing belief in utilitarianism within the scope of discrediting evolutionary explanations after all. As Kahane puts it: "If a disposition to partial altruism was itself selected by evolution, then the epistemic status of its reasoned extension should also be suspect."[26]

Let us grant that utilitarianism represents the reasoned extension of more fundamental evolved evaluative judgments, such as that it is morally right to help one's kin and the members of one's community. Presumably, standard non-consequentialist theories also derive from the same evolved evaluative judgments. Furthermore, it seems plausible that standard non-consequentialist theories hew more closely to these evolved evaluative judgments than do utilitarian moral theories. If this is the case, then, it seems plausible that, to the extent that these

26 Kahane, "Evolutionary Debunking Arguments," 119.

beliefs are debunked, we ought to end up increasing our relative confidence in utilitarianism vis-à-vis other standard moral theories. In other words, we ought to reduce our confidence in standard non-consequentialist theories to a greater extent than we ought to reduce our confidence in utilitarianism, since standard non-consequentialist theories stick closer to the evaluative judgments that end up being debunked. If, in addition, the confidence that we lose in these standard normative theories is redistributed to the hypothesis that nothing matters and so all options are equally choice-worthy, then, since any hypothesis that entails that all options are equally choice-worthy cuts no ice with respect to the appropriateness of the different options available to us under conditions of moral uncertainty, for the reasons explained previously in this paper, it will end up being the case that evolutionary considerations shift our decision making in the direction of utilitarianism by virtue of increasing our confidence in utilitarianism vis-à-vis its standard competitors, even granting that we ought to significantly reduce our confidence in utilitarianism.[27]

4. Conclusion

Assuming that we ought to take normative uncertainty into account, debunking arguments that selectively undermine non-utilitarian theories have genuine practical significance, even if we are also aware of debunking explanations targeting our beliefs about well-being. The latter do not rob utilitarianism of its practical significance. Given the resulting credence distribution over different moral theories and theories of well-being, the most appropriate action will in many cases accord with the action required by utilitarianism in combination with commonsense theories about well-being. Furthermore, the effect of debunking arguments may be similar even if we ought to significantly reduce our confi-

27 It may be objected that it is a mistake to assume that probability mass that is shifted from utilitarianism, deontology, and other normative theories with roots in our evolved moral intuitions should be redistributed to the hypothesis that all options are equally choice-worthy. To the extent that we lose confidence in these different moral theories, it may be argued that we ought instead to gain confidence in *nihilism*, interpreted as the view that all options are *incomparable* in respect of choice-worthiness, as opposed to equally choice-worthy. This need not undermine our argument. Ross argues that nihilism, so understood, can also be ignored under conditions of moral uncertainty ("Rejecting Ethical Deflationism"). MacAskill ("The Infectiousness of Nihilism") raises a number of objections to Ross, but MacAskill, Bykvist, and Ord (*Moral Uncertainty*) go on to outline an improved theory of rational decision-making under moral uncertainty that also allows us to treat all but full confidence in nihilism as practically irrelevant.

dence in utilitarianism in light of evolutionary debunking arguments, so long as other moral theories end up being undermined to an even greater extent.[28]

University of Oxford
andreas.mogensen@philosophy.ox.ac.uk
william.macaskill@philosophy.ox.ac.uk

REFERENCES

Bramble, Ben. "Evolutionary Debunking Arguments and Our Shared Hatred of Pain." *Journal of Ethics and Social Philosophy* 12, no. 1 (September 2017): 94–101.

Brandon, Robert. *Adaptation and Environment*. Princeton: Princeton University Press, 1989.

Crisp, Roger. *Reasons and the Good*. Oxford: Oxford University Press, 2006.

de Lazari-Radek, Katarzyna, and Peter Singer. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press, 2014.

Dennett, Daniel. *Consciousness Explained*. London: Penguin, 1991.

Dutton, Denis. *The Art Instinct: Beauty, Pleasure, and Human Evolution*. Oxford: Oxford University Press, 2009.

Gracely, Edward J. "On the Noncomparability of Judgments Made by Different Ethical Theories." *Metaphilosophy* 27, no. 3 (July 1996): 327–22.

Greene, Joshua. "The Secret Joke of Kant's Soul." In *The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, edited by Walter Sinnott-Armstrong, 35–80. Vol. 3 of *Moral Psychology*. Cambridge, MA: MIT Press, 2008.

Gustafsson, Johan E., and Tom Torpman. "In Defence of My Favourite Theory." *Pacific Philosophical Quarterly* 95, no. 2 (March 2014): 159–74.

Hanson, Louise. "The Real Problem with Evolutionary Debunking Arguments." *Philosophical Quarterly* 67, no. 268 (July 2017): 508–33.

Harman, Elizabeth. "The Irrelevance of Moral Uncertainty." In *Oxford Studies in Metaethics*, vol. 10, edited by Russ Shafer-Landau, 53–79. Oxford: Oxford University Press, 2015.

Jaquet, Francois. "Evolution and Utilitarianism." *Ethical Theory and Moral Practice* 21, no. 5 (November 2018): 1151–61.

Joyce, Richard. *The Evolution of Morality*. Cambridge, MA: MIT Press, 2006.

28 We wish to thank Guy Kahane for comments on an early draft of this paper, as well as the anonymous referees who offered insightful comments and criticisms during the review process.

Kahane, Guy. "Evolution and Impartiality." *Ethics* 124, no. 2 (January 2014): 327–41.

———. "Evolutionary Debunking Arguments." *Noûs* 45, no. 1 (March 2011): 103–25.

Kamm, Frances M. *Morality, Mortality*. Vol. 1, *Death and Whom to Save from It*. Oxford: Oxford University Press, 1993.

Lockhart, Ted. *Moral Uncertainty and Its Consequences*. Oxford: Oxford University Press, 2000.

MacAskill, William. "The Infectiousness of Nihilism." *Ethics* 123 no. 3 (April 2013): 508–20.

———. *Normative Uncertainty*. Doctoral thesis, University of Oxford, 2014.

———. "Normative Uncertainty as a Voting Problem." *Mind* 125, no. 500 (October 2016): 967–1004.

MacAskill, William, Krister Bykvist, and Toby Ord. *Moral Uncertainty*. Oxford: Oxford University Press, 2020.

MacAskill, William, and Toby Ord. "Why Maximise Expected Choiceworthiness?" *Noûs* 54, no. 2 (June 2020): 327–53.

Mogensen, Andreas L. "Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?" *Philosophical Studies* 173, no. 7 (July 2016): 1799–1817.

———. "Evolutionary Debunking Arguments and the Proximate/Ultimate Distinction." *Analysis* 75, no. 2 (April 2015): 196–203.

Nichols, Shaun. "Process Debunking and Ethics." *Ethics* 124, no. 4 (July 2014): 727–49.

Ross, Jacob. "Rejecting Ethical Deflationism." *Ethics* 116, no. 4 (July 2006): 742–68.

Ruse, Michael. *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*. Oxford: Blackwell, 1986.

Sepielli, Andrew. "What to Do When You Don't Know What to Do." In *Oxford Studies in Metaethics*, vol. 4, edited by Russ Shafer-Landau, 5–28. Oxford: Oxford University Press, 2009.

Singer, Peter. "Ethics and Intuitions." *Journal of Ethics* 9, nos. 3–4 (October 2005): 331–52.

———. *The Expanding Circle: Ethics and Sociobiology*. Oxford: Clarendon Press, 1981.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127, no. 1 (January 2006): 109–66.

Tersman, Folke. "The Reliability of Moral Intuitions: A Challenge from Neu-

roscience." *Australasian Journal of Philosophy* 86, no. 3 (September 2008): 389–405.

Vavova, Katia. "Debunking Evolutionary Debunking." In *Oxford Studies in Metaethics*, vol. 9, edited by Russ Shafer-Landau, 76–101. Oxford: Oxford University Press, 2014.

Weatherson, Brian. "Running Risks Morally." *Philosophical Studies* 167, no. 1 (January 2014): 141–63.

White, Roger. "You Just Believe That Because …" *Philosophical Perspectives* 24, no. 1 (December 2010): 573–615.

Wiegman, Isaac. "The Evolution of Retribution: Intuitions Undermined." *Pacific Philosophical Quarterly* 98, no. 2 (June 2017): 193–218.

Wilkins, John S., and Paul E. Griffiths. "Evolutionary Debunking Arguments in Three Domains: Fact, Value, and Religion." In *A New Science of Religion*, edited by Greg Dawes and James Maclaurin, 133–46. London: Routledge, 2013.