
Evolutionary Debunking Arguments and Our Shared Hatred of Pain

Ben Bramble

WHY DO WE THINK IN MORAL AND EVALUATIVE terms (i.e., have moral and evaluative beliefs)? According to some philosophers, it is just because such thinking conferred a fitness advantage on our ancestors (i.e., helped them to survive and reproduce) and we have inherited this disposition. It is not because the things that we morally or evaluatively believe are ever true and we are apprehending or otherwise responding to these truths.¹

In their article, “The Objectivity of Ethics and the Unity of Practical Reason,”² Katarzyna de Lazari-Radek and Peter Singer argue that while many common moral and evaluative beliefs can indeed be evolutionarily debunked in this way – for instance, the belief that incest between consenting adult siblings is morally wrong – at least one belief cannot be so debunked. This is belief in the following principle:

The Principle of Universal Benevolence (UB): each one [of us] is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him.³

The heart of UB is a claim about what is good *simpliciter*: equal amounts of wellbeing are worth equally much, no matter whose they are.

Why can belief in UB not be debunked by evolutionary considerations? It is because, say de Lazari-Radek and Singer, such a belief would have been no assistance to any of our ancestors in the race to survive and reproduce. On the contrary, it would have been an evolutionary hindrance, for it would have inclined one to devote one’s time and resources to promoting the survival and reproduction of beings *other* than oneself and one’s close kin.

Why, then, do we (or at least many of us) believe UB? According to de Lazari-Radek and Singer, a better explanation of this belief is that it is a byproduct of our capacity to reason, which itself was fitness-enhancing. They write:

[We] might have become reasoning beings because that enabled us to solve a variety of problems that would otherwise have hampered our survival, but once

¹ See R. Joyce (2007) *The Evolution of Morality*, Cambridge, MA: MIT Press, ch. 6; S. Street (2006) “A Darwinian Dilemma for Realist Theories of Value,” *Philosophical Studies* 127: 109-66.

² K. de Lazari-Radek and P. Singer (2012) “The Objectivity of Ethics and the Unity of Practical Reason,” *Ethics* 123: 9-31.

³ *Ibid.*: 17. This formulation of the principle is due to H. Sidgwick (1907) *The Methods of Ethics*, 7th ed. (London: Macmillan). De Lazari-Radek and Singer’s piece is organized around two puzzles raised in Sidgwick’s *Methods*. For brevity, I omit discussion here of their interesting claims about Sidgwick’s own views on these matters.

we are capable of reasoning, we may be unable to avoid recognizing and discovering some truths that do not aid our survival.⁴

Having made this case for UB, de Lazari-Radek and Singer move swiftly to a dramatic conclusion: utilitarianism is true. By utilitarianism, they mean the view that there is just one ultimate source of normative reasons for action – *wellbeing, whoever it is* – and so “we ought to maximize well-being generally.”⁵

In his piece, “Evolution and Impartiality,”⁶ Guy Kahane responds to de Lazari-Radek and Singer. According to Kahane, even if de Lazari-Radek and Singer are right that debunking arguments succeed against all beliefs that conflict with belief in UB, but not against belief in UB itself, we can nonetheless evolutionarily debunk *our core beliefs about the nature of wellbeing* – in particular the beliefs that pleasure is good for us and pain is bad for us. It is obvious, Kahane says, that we have these beliefs about the nature of wellbeing only due to their usefulness to our ancestors. He writes: “if *any* normative belief can be given such an explanation, it is the universal (or near-universal) conviction that pain is bad [for us].”⁷

Why is this supposed to pose a problem for the argument of de Lazari-Radek and Singer? It is a problem, Kahane says, because it would mean that the sort of utilitarianism they vindicate is “entirely idle.”⁸ This is because it would be one on which we can have no idea *what* to maximize.

Kahane adds that some beliefs about wellbeing might well survive debunking attempts, but only counterintuitive ones like that wellbeing consists in:

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.⁹

There would have been no evolutionary benefit to having these beliefs. In this case, the sort of utilitarianism de Lazari-Radek and Singer would have vindicated would not be idle, but (worse) it would be one that prescribed perverse actions. Kahane writes:

if we take the goal of purging all evolutionary influence from our normative views seriously enough, we will end up with a view that is so radically divorced from common sense, and so distant from any familiar ethical theory, that, by comparison, Singer’s own utilitarianism will seem almost like old-fashioned common sense.¹⁰

Kahane concludes that utilitarians like de Lazari-Radek and Singer should not invoke evolutionary debunking arguments.

⁴ De Lazari-Radek and Singer (2012: 16).

⁵ Ibid.: 18.

⁶ G. Kahane (2014) “Evolution and Impartiality,” *Ethics* 124: 327-41.

⁷ Kahane (2014: 330-31). For simplicity, in what follows I will focus on the belief that pain is bad for us (though similar points apply to the belief that pleasure is good for us).

⁸ Ibid.: 332.

⁹ Ibid.: 334.

¹⁰ Ibid.: 330.

In what follows, I will respond to Kahane on behalf of de Lazari-Radek and Singer. I will argue that Kahane’s crucial premise (i.e., his claim that a debunking explanation can easily be given of our beliefs that pleasure is good for us and pain is bad for us) is false. On the contrary, these evaluative beliefs are the *hardest* to debunk. The best explanation of them, I will claim, is that we form them as a response to seeing that there is something worth having in pleasure and something worth avoiding in pain. Not only, then, does attention to the origins of our core beliefs about wellbeing not undermine the strategy of de Lazari-Radek and Singer, but it also might vindicate it.

1. The Argument for Kahane’s Premise

Kahane’s key premise is that our core beliefs about wellbeing are “easy targets for evolutionary debunking.”¹¹ He makes this claim several times. He writes:

many of our evaluative beliefs about well-being, including the beliefs that pleasure is good and pain is bad, are some of the most obvious candidates for evolutionary debunking.¹²

And again:

Our evaluative beliefs about pain and pleasure are perhaps the easiest to explain in evolutionary terms. It will be hard, at best, to find a serious evolutionary theorist who would deny this.¹³

And again:

If evolution has had any role in shaping our evaluative beliefs, it would be hard to deny, to put it mildly, that evolution has played a key role in disposing us to see pain as bad, as something we have strong reason to avoid and minimize.¹⁴

What is Kahane’s argument for this claim? The only argument he offers is one given by Sharon Street. He cites her as follows:

it is of course no mystery whatsoever, from an evolutionary point of view, why we and the other animals came to take the sensations associated with bodily conditions such as [cuts, burns, bruises, broken bones] to count in favor of what would avoid, lessen, or stop them rather than in favor of what would bring about and intensify them. One need only imagine the reproductive prospects of a creature who relished and sought after the sensations of its bones breaking and its tissues tearing; just think how many descendants such a creature would leave in comparison to those who happened to abhor and avoid such sensations.¹⁵

¹¹ Ibid.

¹² Ibid.

¹³ Ibid.: 334. Kahane does not cite any evolutionary theorists.

¹⁴ Ibid.: 331.

¹⁵ Street (2006: 150).

It is unclear exactly what the argument here is supposed to be. Of course, a being who *relished and sought after* the sensations of its bones breaking and tissues tearing would not be very fit, evolutionarily speaking. But how does this tend to show that *believing that one's pain is bad for one* would have been fitness-enhancing? That wanting and seeking one's pain makes it more likely one will be injured has no obvious tendency to show that believing that one's pain is bad for one makes it less likely that one will be injured.

Let us, then, ignore Street's claim about relishing and seeking after pain, and ask directly: does believing that one's pain is bad for one make it more likely that one will avoid injury?

The answer, it seems to me, is that it would do so *only if one does not already hate or have an aversion to one's own pain*. If one already hates one's own pain, then one already has a considerable motive to avoid such pain. Coming to additionally believe that one's pain is bad for one would seem to add little or nothing to one's tendency to avoid such feelings, and so to one's evolutionary fitness. Indeed, in creatures who already hate their own pain, adding a belief in the badness for them of pain might, at least in certain circumstances, *reduce* their fitness by leading them to focus excessively on immediate pain relief to the exclusion of more important things (such as the survival or flourishing of kin or certain long-term interests of their own).

Suppose this is right. A revised debunking argument is nonetheless available to Kahane/Street. According to this argument, while the belief that one's pain is bad for one is not *itself* fitness-enhancing, we have such a belief *only as a result of our hatred of pain*, and we have our hatred of pain only because such hatred was fitness-enhancing in our ancestors. If this is true, then there does appear to be an effective evolutionary debunking argument against the belief that one's own pain is bad for one. Why do we have this belief? Merely as a byproduct of a different state that itself is something we have only for its contribution to our fitness. This, indeed, may be the argument that Kahane/Street had intended to give all along, but failed to put clearly.

2. Evaluating the Argument for Kahane's Premise

What should we make of this revised argument for Kahane's premise? I will not consider the part of it that says that our belief that pain is bad for us is the result of our hatred of pain (though clearly this claim itself needs much fleshing out if the argument is to succeed). I want, instead, to focus on the part of the argument that says that our hatred of pain is due to the usefulness of such hatred in our ancestors. Would such hatred have assisted our ancestors in avoiding injury and so surviving and reproducing?

I think it unlikely that we can explain our hatred of pain in this way. It is certainly true that hating pain *nowadays* (and, more generally, at most times in our recent evolutionary history) is useful in avoiding injury. Young children, especially, who happen not to hate the feeling of pain certainly have a tendency to get injured and die. This is because pain is correlated with injury.

But *why* is pain correlated with injury? Why, in other words, is injury painful in the first place? Kahane and Street seem to assume that it is simply a brute fact that injury is painful. But this might not be so. Perhaps bodily injury was not painful in the beginning at all (i.e., in early beings from whom we are descended), but only *came* to be painful for beings like us as a result of evolution's selecting for beings for whom injury merely happened (i.e., as a result of random mutation) to be painful. Why would evolution have selected for such beings? Perhaps because *all beings start out with a hatred of their own pain*. If all beings start out with a hatred of their own pain, then beings who merely happen to feel pain when they are injured will naturally do much better than their rivals when it comes to avoiding such injury.

On this account, it was hatred of pain (not the painfulness of injury) that was a given, and was what evolution made use of to get beings to more effectively stay away from things that could injure them. Evolution, in other words, did not make us hate pain by weeding out those who happened to be indifferent to or like it. Instead, it made injury painful by weeding out those for whom it happened not to be painful.

Suppose this is true, that bodily damage is painful for us today only because our distant ancestors had a preexisting hatred of their own feelings of pain. The vital question now is: what explains this preexisting hatred?

The realist has an intriguing answer. She can say that our hatred of pain is in some way *a response to pain's objective badness for us*. Perhaps the badness for one of pain is a normative fact that is so obvious that any creature, even quite basic creatures, can apprehend it (indeed, cannot *avoid* apprehending it). We see that pain is bad for us, and this is why we are all inclined to avoid it.¹⁶ Alternatively, there is some way in which pain's objective badness for us leads us all to hate pain *without* it being necessary that we realize or apprehend its badness for us.¹⁷

3. A Third Possibility

I have outlined two possible orders of explanation. First, there is that of Kahane/Street, on which the connection between pain and injury is basic, and we have evolved a hatred of pain because such hatred helped our ancestors avoid injury. Second, there is the order of explanation I have proposed, on which it is our hatred of pain that is basic (and in some way a response to pain's objective badness for us), and evolution selected for beings for whom injury merely happened to be painful, thereby leading to injury being painful for all of us today.

¹⁶ As Irwin Goldstein says in connection with this issue, "creatures of elementary intelligence are still capable of elementary insights." See I. Goldstein (1980) "Why People Prefer Pleasure to Pain," *Philosophy* 55: 357. The suggestion here is not that these animals can necessarily think about value, or have evaluative thoughts in anything like a language like our own, but just that they can be aware on some level, and in some way, that when they are in pain something bad is going on.

¹⁷ A further possibility, it must be conceded, is that there is an explanation of our preexisting hatred of pain that appeals neither to its usefulness in evolution nor to its objective badness for us. I will not further consider this possibility here.

But there is a third possibility I have not yet mentioned. This is that neither the connection between pain and injury nor our hatred of pain is basic. Instead, these things evolved *together, as jointly advantageous*.¹⁸ What was fitness-enhancing was hating *something* that was correlated with many different ways in which one might get injured or have one's fitness reduced. Such hatred of a single object (in this case, a feeling produced by each of these threats to one's fitness) was more efficient, or "cheaper" in evolutionary terms, than evolving independent desires to avoid each of these different threats. If this third possibility were true, then we would have a successful evolutionary debunking explanation of our belief that pain is bad for us.

What should we make of this hypothesis? I cannot fully address it here, but I believe it is implausible for the following reason. If the sole function of the feeling of pain were to be a feeling we hate, then it seems, at least theoretically, that the feeling of pain could have been *any old feeling*. Any feeling at all might equally well have done the job. But this, intuitively, is not what we find to be the case. The feeling of pain is qualitatively unlike any other feeling, in the following respect: it is one that seems to be in some sense *worth* hating or being averse to. Its aversiveness seems somehow to be built into its very nature.

Now, you might object that this is exactly how we should expect to feel about a feeling that was selected for us to be averse to it. We should, of course, expect to find it aversive. But there is a difference between finding something aversive in the sense merely of wanting it not to occur (the only sense in which, on this third possibility, it should be necessary that we are averse to the feeling selected to be pain), and finding it aversive in the sense in which we find actual pain aversive. Many things we hate, or want not to be the case, are not aversive to us in remotely the way that pain is aversive to us. Why could the feeling of pain not have had a quality more like one of these other things? The peculiar and unique phenomenal quality of pain, I am suggesting, gives us some reason to think that it was not chosen merely as a "thing for us to hate," but that we hate it in some sense because it merits our hatred.

4. Conclusion

What is the upshot? Not only does attention to the origins of our core beliefs about wellbeing not undermine the strategy of de Lazari-Radek and Singer, it may well vindicate it. The beliefs that pleasure is good for us and pain is bad for us may have a firmer foundation than any other evaluative beliefs we have – a very firm foundation indeed.

De Lazari-Radek and Singer claim that the capacity to reason helped us to evolve and then, *as a side effect*, enabled us to recognize normative truths, like that of utilitarianism. If I am right, however, our ability to recognize normative truths is not only a side effect of the usefulness of our capacity to reason, but also *an essential part of its usefulness*. It was our ability to recognize the goodness of pleasure and the badness of pain in the first place that got us wanting

¹⁸ I am grateful here to an anonymous reviewer for this journal.

pleasure and hating pain, the very attitudes that evolution then exploited to direct us toward food and away from bodily injury.

Ben Bramble
Department of Philosophy
Trinity College Dublin
brambleb@tcd.ie