Does suffering *really* matter?

Nicolas Delon

New College of Florida

Abstract. Is suffering really bad? In a robustly realistic sense? In *On What Matters*, the late Parfit argued that we all have reasons to want to avoid future agony and that the suffering is in itself "doubly bad", for the one who suffers and impersonally, and it is so objectively and intrinsically. Call it Realism about the Value of Suffering (RVS). This paper has two aims. It argues against RVS by drawing from a broadly genealogical debunking of our evaluative attitudes, showing in particular that each of the responses to debunking available to the objectivist fail. It also argues that a Nietzschean approach is well suited to support the challenge, bearing in mind that, for Parfit, Nietzsche is the ultimate test of his hope that we can converge on fundamental normative truths. Here, I conclude that whether suffering matters is really a matter of our attitudes rather than a mind-independent normative truth.

Keywords: suffering, pain, value, realism, debunking

Total word count: 8,590

1. Introduction

According to moral realism, moral facts are mind-independent. One such fact could be the nearly universally held belief that suffering doesn't just feel bad, but is in itself bad. Call *Realism about the Value of Suffering* (RVS), both physical and psychological (including but not exclusively pains) the view that

> In all possible worlds, suffering is bad regardless of our evolutionary history and of how our attitudes to it could differ

On RVS, there is such a thing as the *correct* attitude to suffering—one's own or someone else's—namely, judging that it is bad and ought not to occur or continue. The judgment is true whether or not we consider that the badness of (say) a headache consists in how it feels. The felt character of pain may constitute or simply coincide with its badness. Regardless, RVS states, it is bad.

Suffering occupies a central place in the work of two philosophers who could hardly be more different: the late Derek Parfit, from *Reasons and Persons* (1984) to *On What Matters* (2011; 2017) and Nietzsche throughout his published and unpublished work. In the second volume of *On What Matters,* Parfit dedicates a whole chapter to Nietzsche, right after a discussion of moral disagreement and convergence on normative truths,[1] and describes him as "the most influential and admired philosopher of the last two centuries" (OWM II 570; cf. 1984: 176).

Parfit argues for *Double Badness of Suffering* (DBS):

> All suffering is in itself both bad for the sufferer and impersonally bad (OWM II 569)

The badness of suffering is the canonical example of an irreducibly normative truth: we have reasons, personal and impersonal, to want pain to cease or not occur, and such reasons are primitive: they do not depend on our attitudes or other facts. Truths about suffering are not explained by our desires. Rather, our desires are justified only in so far as they track our objective reasons. The disvalue of pain (mine, yours, anyone's) is independent of anyone's attitudes.

Nietzsche denies not just DBS but that any suffering is *in itself* bad. According to Parfit, on the other hand, suffering is intrinsically bad and all sufferings matter equally. But Nietzsche denies both claims. It follows that Nietzsche denies Parfit's account of our reasons to reduce suffering, for it would not, he thinks, make the world or anyone's life necessarily better. Nietzsche therefore appears to threaten Parfit's hope for normative convergence. We cannot ignore Nietzsche, Parfit claims.

Is RVS susceptible to genealogical debunking? Or could it be that, unlike many moral beliefs, our attitudes to suffering are immune, especially when we note the nearly universal consensus and explain

[1] Hereafter I will use the abbreviation OWM followed by volume and page (e.g., OWM II 571). I will use the conventional abbreviations for Nietzsche's works (see below for cited works), cited by part and section (e.g., G: II: 3). Posthumous fragments are cited in accordance to their classification in the *Kritische Studienausgabe*, with volume and page numbers when applicable.

A	*The Anti-Christ*
BGE	*Beyond Good and Evil*
D	*Daybreak*
GM	*On the Genealogy of Morality*
GS	*The Gay Science*
KSA	*Kritische Studienausgabe.*
TI	*The Twilight of Idols*

away the remaining disagreements as the product of distorting influences (as Parfit purports to)? Parfit is not alone in defending pain's intrinsic badness, whether it be to vindicate hedonism about wellbeing, objectivism and/or utilitarianism. These defences need not presuppose non-naturalist realism *à la* Parfit, but they do involve some version of RVS. This paper has two aims. I argue that these defences fail to vindicate the attitude-independent disvalue of suffering, and that a broadly Nietzschean approach brings to focus the core of the challenge. Why? For one thing, Parfit's canonical example of a non-debunkable belief is tried against Nietzsche's contrarian views. Further, there are obvious, if imperfect, parallels between contemporary evolutionary debunking arguments and Nietzsche's genealogy. Finally, some aspects of Nietzsche's psychology that find echo in recent empirical psychology and cognitive science are precisely ones that could explain our attitudes to suffering without presupposing RVS. The remainder of this paper proceeds as follows. Sections 2-4 set the stage and summarize Parfit's case for RVS—his metaethics, his argument for normative convergence, and his conception of the disvalue of suffering. Sections 5-9 motivate the debunking challenge against RVS and draw connections between recent empirical work and the import of Nietzschean hypothesis regarding the genealogy of our attitudes to suffering.

2. Parfit's metaethics

In OWM, Parfit argues that there is a unique moral principle, on which the most plausible moral theories (utilitarianism, Kantianism and contractualism) converge (the "Triple Theory"). But, thinks Parfit, for convergence to be significant, it must be the case that moral claims are objectively true and irreducibly normative. Case in point, normative claims about suffering are made true by objective moral facts that are modally necessary and causally and constitutively independent of our attitudes. Parfit seeks to vindicate cognitivism (moral judgments express propositions and are truth-apt), realism (moral truths are independent of our judgments and attitudes), non-naturalism (moral facts are not identical to or reducible to natural facts) and externalism (what reasons we have does not depend on our motives). In sum, moral facts are mind-independent irreducible normative truths on which we can converge (OWM II 544). These truths are not a matter of empirical discovery but known by intuition and rational argument. (OWM III 200-1)

Does metaethics have anything to say about whether we should relieve my (say) headaches (and yours, anyone's) or care about the horrors that afflict strangers and other fellow creatures? In his review of OWM, Allan Gibbard (2012) writes that, "[s]uffering matters whatever the story of normative properties turns out to be". Yet to Parfit the story matters greatly. It matters that we have objective reasons to care about everyone's well-being, grounded in the intrinsic badness of suffering, because this belief can lead us to do more to prevent or relieve their suffering. (OWM III 190) More generally, if it turned out naturalism were true (i.e., if there were no irreducibly normative facts), "Sidgwick, Ross, I, and others would have wasted much of our lives. We have asked what matters, which acts are right or wrong, and what we have reasons to want, and to do." (OWM II 367)[2] And for things to really matter, it must be possible to be objectively right about it. Hence, the need to vindicate non-naturalism and eliminate disagreement. Our *really* having normative reasons hinges on there being irreducible normative truths. It would be "a tragedy" if there were no moral facts or if we could not agree on a single true morality.

Simon Blackburn (2011) quipped that "outside the charmed walls of All Souls College, there actually are tragedies." And Max Hayward (2018: 735) notes the following irony:

> the truth can be uncomfortable and even unpleasant. … Realists, and especially non-naturalists, have long argued that truth in ethics is independent of whatever anyone happens to think about it: to call something 'true' is to do something *over and above* endorsing, recommending, or approving of it.

The truth of naturalism could be tragic from Parfit's standpoint, but Hayward rightly notes it doesn't "count as evidence against its truth." Still, one may want to show that suffering *really* matters in some robust sense. Gibbard may not think that RVS bears on whether "suffering matters," but for Parfit it does, and I will assume that he is right to worry that, if RVS is unjustified, our reasons to reduce suffering may not be unequivocal or often sufficient.

[2] Naturalists such as Peter Railton (2017) of course disagree that they cannot countenance normative claims about what we have "most reason to do", including about the significance of pain: moral concepts specify the "job descriptions," fulfilled by natural properties, featured in our substantive moral claims. Most contributors to Singer (2017) seek to deflate Parfit's tragic implication that nothing could matter if naturalism were true.

3. Convergence and disagreement

DBS is among the judgments on which everyone would ideally agree. In fact, nearly universal agreement already exists about the badness of *undeserved* suffering. Parfit hopes that under ideal conditions we will someday all come to recognize normative truths regardless of our metaethical disagreements (OWM II 550). These conditions are described by an empirical prediction, the *Convergence Claim* (CC):

> If everyone knew all of the relevant non-normative facts, used the same normative concepts, understood and carefully reflected on the relevant arguments, and was not affected by any distorting influence, we and others would have similar normative beliefs. (OWM II 546)

Because we can all make mistakes, even under ideal conditions, Parfit adds that "it would be enough to defend the prediction that, in ideal conditions, we would *nearly* all have *sufficiently similar* normative beliefs." (p. 547)

CC specifies the conditions under which reasonable disagreement takes place and can be resolved. If Parfit can show that Nietzsche does not meet the above conditions (e.g., he lacks a proper concept of reasons, he engaged in motivated reasoning, or his mind deteriorated) then he can show that Nietzsche's views do not constitute an objection to CC. But first, what is Parfit's argument for RVS?

4. Reasons of Agony

According to the *Argument from Agony* (OWM I, ch. 3-4), suffering provides irreducibly normative reasons. If so, suffering constitutes a counter-example to Parfit's meta-ethical rivals, for anti-realists and naturalists cannot explain this basic fact (OWM II 551). RVS is a normative truth in the "reason-implying sense": we necessarily have reasons to want to avoid future suffering, which we know intuitively just as we know it's wrong to torture people for fun. RVS is not a psychological claim: "Naturalists could point out that nearly everyone does want to avoid future agony, at least if this agony would be in the fairly near future. But the fact that people *have* this desire cannot give them a *reason* to have this desire." (OWM III 67)

Parfit does not argue that suffering is universally psychologically aversive.[3][4] Our actual desires and preferences do not bear on normative truth. It is the value of suffering that explains our dislike, and which of our desires are rational, not the other way around. Many anti-realists concede the premise that we all have reasons to avoid future agony, but they deny that it follows from it that our reasons are objective (i.e., subjectivism *can* explain our attitudes; cf. Sobel 2011). But Parfit argues that subjectivism, by grounding our reasons on our desires, makes them depend on desires which we have no prior reasons to have—a normative quicksand, as it were (OWM I 73-82). Still, subjectivists and Parfit could agree that suffering is always bad, even if the former denied RVS robustly construed. On the other hand, Nietzsche does not simply reject the conditional that, if we all have reasons to avoid agony, such reasons must be objective. He denies the antecedent.

In the context of Parfit's argument for normative convergence, the *Argument from Agony* implies that the axiological status of suffering bears on our normative reasons. Suffering is a source of impartial, or "person-neutral", reasons to reduce it in so far as no reference to me or anyone in particular figures into the description of the relevant considerations. We all have reasons to regret, prevent or relieve the suffering of anyone (OWM I 138; cf. Nagel 1986: 161: "the pain can be detached in thought from the fact that it is mine without losing any of its dreadfulness. … suffering is a bad thing, period, and not just for the sufferer.") Hence, the stakes are high. For Parfit's normative views hinge on his success establishing a robust form of moral realism. The disvalue of pain serves as a key bridging principle.

Now that the stage is set and the stakes spelled out, I set out to reconstruct the Nietzschean debunking challenge.

[3] Most of Parfit's examples involve ordinary physical pains like being whipped or burnt, like the sensation of red-hot iron (OWM II 459, 541), but he recognizes that psychic suffering can be worse (p. 569).

[4] There are two main theories of pleasure and pain: the *phenomenological* theory and the *attitudinal* theory. On the former, a pain or pleasure is defined by how it is felt; on the latter, it is defined by its relation to the subject's pro-/con-attitudes towards the sensation. Eden Lin has recently argued for a hybrid theory (forthcoming). Parfit also appears to endorse a mixed view: it is not the sensations that are bad but *our awareness of having a sensation that we dislike* (OWM I 54), i.e., the hedonic character of conscious states. Since Parfit is concerned with the *axiology* of suffering, however, little turns on this question. Objective reasons must be independent of our desires *and* of the phenomenology of suffering. Further, that our desires are justified by the nature of suffering is not a conceptual but a *normative* truth (OWM II 551). Parfit thus seems to allow for a dissociation between certain sufferings (say, a migraine) and the reasons to avoid them: you would be evaluatively mistaken to value the migraine.

5. Motivating the debunking challenge

Generally, evolutionary explanations of our moral beliefs seek to undermine moral objectivism by casting our beliefs as systematically prone to error, modally contingent or better explained by non-truth-tracking processes. This is not to say that our beliefs are false; merely that our justifications for holding them are undercut in the absence of countervailing evidence. Synchronically, to the extent that our evaluative beliefs depend on "epistemically defective emotional and motivational processes" they are unjustified. (Nichols 2014) Diachronically, our moral judgments and norms can be traced to their cultural and environmental origins in our hominid ancestors. The explanation of how we can to moralize our responses to violations of evolved prosocial norms is only contingently related to what we take to be the moral truth. Had our evolution been different, we might have ended up with radically different beliefs, and contingency poses a problem because moral discourse presupposes invariant moral reasons (Joyce 2001; 2016). If we further assume that the evolutionary processes that gave rise to our moral beliefs are unreliable, then even minor variations along the evolutionary path would have led us to mostly false moral beliefs, regardless of what the moral truths are (Street 2006).

The history of morality is rife with radical variations in moral norms across time and cultures. Nichols (2004), for example, documents how "harm-norms" have gradually evolved to expand the scope of moral concern (alongside both an enhanced sensitivity and lessened exposure to the suffering of others) to include most or all human beings and even nonhuman animals. We presently disapprove of many moral norms (e.g., cruel punishments or a greater disposition to violence) that our ancestors widely accepted. Cultural evolution, it turns out, accounts pretty well for the stability *and* the disappearance of such norms. Objectivism doesn't. Our best explanation might even be an error theory (Joyce 2001; Mackie 1977).

In sum, debunking challenges argue that causally plausible but epistemically unreliable processes—to which we do have epistemic access—undercut our justifications for our beliefs.[5]

Parfit views moral beliefs as immune to evolutionary debunking because they are discovered by reason, like logical, and mathematical truths. To the extent that our epistemic reasons are immune to

[5] The evolutionary challenge does not establish that *no* realist explanation is forthcoming, simply that the realist cannot offer an evolutionary or trivial explanation of why we would have many true moral beliefs (Clarke-Doane 2012).

evolutionary debunking, we can assume that we are able to discover at least some irreducibly normative practical reasons.[6] One of his main interlocutors is Sharon Street. She argues that, since our "basic evaluative tendencies" were shaped by evolutionary forces unrelated to the moral truth, most of moral judgments, for all we know, may be off-track. The realist must show that an incredible (and inexplicable) "coincidence took place—claiming that as a matter of sheer luck, evolutionary pressures affected our evaluative attitudes in such a way that they just happened to land on … the true normative views." (Street 2008: 208-9) But Street argues that the truth of our beliefs need not figure into the explanation of why we hold the beliefs we do. If we have a complete evolutionary (more broadly, naturalistic) explanation of our beliefs, and we were not selected to have true moral beliefs, then we would most likely have evolved the same beliefs *even if* the moral truth turned out to be wildly different from what we believe. We need not even assume that moral facts could be different; we would have ended up with the same beliefs regardless, given that the underlying causal processes are off-track.

Parfit (2011), among others (e.g., Lazari-Radek and Singer 2013), responds that rational reflection exerts a countervailing pressure on the formation of our moral beliefs which makes them reliable. Can this strategy work? Street had already denied that reason could put us on track (2006: 124). The rationalist objection, she writes, assumes that

> rational reflection provides some means of standing apart from our evaluative judgements, sorting through them, and gradually separating out the true ones from the false as if with the aid of some uncontaminated tool. … [But] [i]f the fund of evaluative judgements with which human reflection began was thoroughly contaminated with illegitimate influence … then the tools of rational reflection were equally contaminated.

Lazari-Radek and Singer reply that we have reasons to trust, not our starting points, but the faculty of reasoning itself. Plus, the fact that it was selected for is irrelevant to the contents of the beliefs it produces. Once we evolved a capacity for rational reflection, our ability to discern objective moral truths came with a package that could not be "economically divided" (2017: 288). Street may be right that our starting points could be contaminated, but objectivists point to a reliable process. Yet, as

[6] Can Parfit avail himself of this strategy? The question is beyond the scope of this paper, but for an argument that accepting the upshot of the evolutionary debunking of moral realism commits one to a similar debunking of the objectivity of mathematical truths, see Clarke-Doane (2012).

Hayward aptly puts it, "Street's point is that rational reflection cannot turn muck into gold." (2018: 726) Given contaminated starting points, rational reflection is no more than "a process of assessing evaluative judgements that are mostly off the mark in terms of others that are mostly off the mark." (Street, ibid.) Positing reason, then, will not suffice to insulate our beliefs.[7]

I will return to what may ground this response to evolutionary debunking. In what follows, I sketch recent responses to the broad debunking challenge, illustrated by four objections (peer disagreement, genealogy, opacity, interpretation), and argue that each fails to immunize RVS.

6. The Objection from Peer-Disagreement

Most evolutionary debunking arguments appeal to the causal processes underlying the formation of our beliefs, but evolutionary challenges can also be drawn from disagreement. According to the *Argument from Counterfactual Peer-Disagreement*, because in evolutionary scenarios our counterparts probably have beliefs incompatible with ours, our beliefs are epistemically contingent, and contingency contradicts the objectivity (i.e., necessity) of our beliefs.[8] The strategy purports to be "justification-defeating." Michael Klenk (2018) argues that it fails because the disagreement it reveals is *not epistemically significant*. Counterfactual disagreement occurs in scenarios that are either nearby or distant. In nearby scenarios, disagreement does not affect most of our beliefs; in distant ones, disagreement is significant but does not occur amongst epistemic peers. Certain moral beliefs would thus survive the challenge.

---

[7] Skarsaune (2011) offers a conditional argument according to which, *if* pain is bad, then we can offer a causal story of why we reliably came to form justified beliefs about its badness. But note that he does *not* offer any argument as to whether pain is indeed bad. For doubts about this strategy when it comes to defending something like RVS against Street's debunking, see Kahane (2014).

[8] This argument is distinct from the classic argument from disagreement against moral objectivity (e.g., Mackie 1977), which has two main steps. First, there is no significant convergence on fundamental moral beliefs. Unlike the natural and social sciences, moral theory has made little progress. Second, the best explanation for such disagreements is that moral facts are not objective (or that there are no moral facts at all). Hence, if we have epistemic reasons to prefer theories that have more explanatory power, we should prefer anti-realism (OWM II 542, 546; also see Leiter 2014). Parfit takes issue with both steps of the argument. Through the history of morality, "[w]e find a series of challenges to established beliefs, which lead to plausible revisions, and to greater agreement. … [T]here has been slow but accelerating progress towards the beliefs that everyone's well-being matters equally, and that everyone has equal moral claims. (OWM II 563) Early distortions caused by evolutionary forces "are being erased, such so that more and more people have true moral beliefs." (p. 538)

If Klenk is right, it could be that disagreement between Nietzsche and Parfit about DBS is not significant either because they do not in fact disagree or because they are not epistemic peers. As recently argued by two commentators, the disagreement between Parfit and Nietzsche is real and significant (Huddleston 2017; Janaway 2016;). Nietzsche did claim, coherently, that suffering could have positive value, both for oneself as a matter of flourishing and for culture as a matter of fostering greatness. Nietzsche, that is, denies DBS. In fact, Parfit and Nietzsche hardly share any fundamental moral belief. For instance, Nietzsche and Parfit seem to agree that we have reasons to promote flourishing. But the agreement is shallow and ends up undermining Parfit's attempt to debunk Nietzsche. For, if Nietzsche does possess the relevant concept of reasons, then his reasons stand in stark contrast to the reasons Parfit claims we have to promote well-being. Their respective views of well-being and of its normative implications diverge radically. Parfit claims: "On all plausible theories, everyone's well-being consists at least in part in being happy, and not suffering" (OWM I 101), and all plausible moral theories agree that everyone's wellbeing matters equally. Nietzsche, on the other hand, argues that suffering can be good and that the well-being of all does not matter for its own sake or equally (cf. Huddleston 2017).

How about the second response? Taking Parfit at his word, Nietzsche should be considered an epistemic peer. After all, he is the ultimate test of CC. One might, like Parfit does, attribute the remaining disagreements to distorting influences. Then Nietzsche would only be an epistemic peer when they do *not* disagree, but we just saw there are no such cases. So, this response fails too. For Parfit's argument that Nietzsche does not reason under epistemically appropriate conditions is far from convincing (Huddleston 2017; Janaway 2016). We are then left with genuine peer-disagreement about fundamental moral beliefs. Hence, Klenk's response to the *Argument from Counterfactual Peer-Disagreement* is only available to Parfit if we assume that Nietzsche is not an epistemic peer. Since we cannot assume so without begging the question or opening Parfit to a symmetric challenge, the argument fails to show that beliefs about suffering are immune to debunking.

7. Genealogical Debunking

Now let's zero in on the genealogical challenge to RVS. Specifically, how does genealogy cast doubt on DBS?

First, genealogy can uncover the detrimental effects of some of our values—not just their suspicious origins ('*pudenda origo*') but their nefarious consequences. Joyce captures it nicely:

> [Genealogy] can reveal that the circumstances that rendered … affective states adaptive on the African savannah (say) no longer hold in the modern world, or fail to hold in some particular circumstances. Genealogical evidence can act as a defeater of the benefit of the doubt we would otherwise accord an affective state—over-turning the assumption of its contribution to our welfare. (Joyce 2016: 173-4)

Many Nietzschean hypotheses have found empirical support confirming the affective and largely subconscious causes of our judgments and decisions as well as our tendency to motivated reasoning, confabulation, and *post-hoc* rationalization (Knobe and Leiter 2007; Telech and Leiter 2016; cf. D 34, GS 335, BGE 5, KSA 13:14 [116]; Haidt 2000; Greene 2007; 2013). His hypotheses, alongside evolutionary arguments, put further pressure on realism. Nietzsche's diagnosis of our sensitivity to suffering appears to fulfil this critical role by tracing our attitudes to their physiological causes and explaining morality as a "sign-language of the affects" (BGE 187). For our sensitivity admits of a psycho-physiological diagnosis and is contingent and historically recent (cf. BGE and GM; Nichols 2004, ch. 6-7). To the extent that Nietzsche's diagnosis applies to morality, it applies to Parfit: RVS might be a symptom of a hypersensitivity to suffering that itself is explained by historical and sociological facts. As Nietzsche puts it provocatively, the morality of compassion is "just another expression of … physiological overexcitability" (TI "Skirmishes": 37).

On such a view, the evolution of harm-norms explains the rise of (say) hedonism and utilitarianism more than the other way around. "Nietzsche might just as well claim," writes Huddleston, "that the people who denounce suffering as always in itself bad are equally beset by a serious form of psychological distortion. Their weakness and 'softening' [TI "Skirmishes": 37] make them fetishize the phenomenal character of suffering" (2017: 180; also see GS 48; BGE 225). According to a

Nietzschean diagnosis, thus, Parfit is susceptible to distorting influences of his own and cannot be brought to appreciate the intelligibility of Nietzsche's view for the pervasive influence of morality.

Admittedly, Parfit was aware of the limited strength of genealogical objections:

> if some attitude has an evolutionary explanation, this fact is neutral. It neither supports nor undermines the claim that this attitude is justified. But there is one exception. It may be claimed that, since we all have this attitude, this is a ground for thinking it justified. This claim is undermined by the evolutionary explanation. Since there is this explanation, we would all have this attitude even if it was not justified; so the fact that we have this attitude cannot be a reason for thinking it justified. (1984: 186).

Again, in *On What Matters*:

> Nietzsche makes some fascinating claims about the origins of morality, especially Christian morality, and he sometimes suggests that these claims undermine morality. But as Nietzsche himself points out, that is not so. When we learn about the origins of morality, or of many other features of human life, we learn very little about the present state, or value, of these things. In Nietzsche's words, 'The more insight we possess into an origin the less significant does the origin appear.' [D 44] (OWM II 583)[9]

Yet genealogy can undercut or defeat the epistemic justifications of our beliefs, as we saw. Parfit confuses the fact that genealogy cannot on its own falsify our beliefs with the fact that it does not bear, epistemically, on their justification. But any genealogy, to the extent that it is plausible, leaves the question of whether a given belief is justified open. Unfortunately, Parfit falls short of confronting genealogy directly and appears to treat this open question as (incorrectly) implying the normative insignificance of genealogy.

[9] Also see GM II 12; GS 345; KSA 12:2 [131] and [189]. Note that D 44 does *not* concern genealogy. Nietzsche means that the origin lacks authority. Parfit seems to have in mind something like the genetic fallacy, which Nietzsche averts.

Remember that Parfit is attempting to debunk Nietzsche's views. However, Huddleston notes, he risks throwing stones from a glass house, an objection also raised by Kahane to evolutionary debunking arguments meant to vindicate utilitarianism (e.g., Singer 2005; Greene 2007):

> utilitarianism is empty of content unless supplemented by an account of well-being. But 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.
>
> Why should we expect that, if some evaluative beliefs can survive the doxastic purge, the resulting normative view would resemble any of the present competing alternatives? After all, all of these, including utilitarianism, were developed by reflection on a set of evaluative beliefs and intuitions that is at least partly infected by distorting influence. … [I]f anything would survive [a purge of all evolutionary influences], it is likely to be far more counterintuitive than anything dreamed of by utilitarians. Perhaps we would need to reject the very normativity of well-being, or at least replace our current attitudes to pleasure, pain, health and death with an especially elevated form of perfectionism. These are only speculations, but, worrisomely, the view that emerges in outline is more Nietzsche than Singer. (Kahane 2011: 120)[10]

According to Street (2006: 150), there is a ready explanation for such attitudes:

> 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.

Such attitudes enhanced our fitness. And Kahane's main claim is that debunking cannot be tailored to support utilitarianism without threatening to debunk our views about well-being themselves. It is a risky strategy. To the extent that Parfit employs debunking, he risks undermining the foundations of his own edifice.

[10] Also see Kahane (2014: 330-2). Kahane (2016), however, cautiously holds the (nearly universally held) view that pain is inherently bad. His reply to the evolutionary challenge is similar to those reviewed in section 8 below, but his 2011 and 2014 articles clearly express doubts about the realist strategy.

Again, the processes that lead to our attitudes are not truth-tracking. However, evolutionary debunking does not deny the possibility of other explanations. One possible explanation appeals to our possessing a certain sort of reliable faculty. Yet, as I argue in the next section, it falls prey to another objection from epistemically defective processes.

8. The Objection from Opacity

Parfit argues that beliefs like that the well-being of all matters equally could not have been reproductively advantageous; instead, we have them because, as rational beings, we were able to recognize their truth (OWM III 340; cf. Lazari-Radek and Singer 2013; 2017). But note that even when the faculty appealed to is reason, one must presuppose some basic fact that is intuitively accessed for any rational argument to gain traction. Fundamentally, the strategy relies on the reliability of introspective access to the badness of suffering. So, I now turn to two attempts to ground such access in response to debunking challenges.

According to Ben Bramble, Street's explanation does not show that *believing* that pain is bad increases our fitness; the belief is superimposed on a prior and sufficiently motivating *aversion* to pain (Bramble 2017: 97; cf. OWM II 527-8). Such beliefs, he argues, are "*the hardest* to debunk." (p. 96) Their best explanation is our ability to perceive the intrinsic axiological quality of pain and pleasure. The fact that our physiological dispositions admit of evolutionary explanations is irrelevant to the truth of our beliefs, whereas the reason we came to form such beliefs does have to do with their truth. Precisely because we were already sufficiently motivated, we could not have developed such beliefs by way of evolutionary adaptations. Rather, our beliefs are best explained by the truth to which they gave access to the cognitive creatures that we are.

Similarly, Neil Sinhababu (manuscript; cf. Lazari-Radek and Singer 2013: 267) argues that we can discover the badness of pain by simply experiencing it: "Phenomenal introspection, a reliable way of forming true beliefs about our experiences, produces the belief that pleasure is good" (p. 18[11]) (or pain bad). Sinhababu assumes the reliability of phenomenal introspection and goes on to argue that our

[11] I refer to page numbers in the typescript available at https://philpapers.org/archive/SINTEA-3.pdf

intuitive judgments about the badness (goodness) of our pain (pleasure) carry over to impartial judgments: "Even though the only pleasure I can introspect is mine now, I should believe that others' pleasures and my pleasures at other times are good … My argument thus favors the kind of universal hedonism that supports utilitarianism." (p. 23)

Sinhababu's argument is an answer to the sort of arguments marshalled by Joyce and Street. On his view, phenomenal introspection is

> a process of belief-formation that evolved to be generally reliable, like visual perception. … [C]reatures who could reliably form true beliefs about their phenomenal states would be more likely to survive and reproduce. Hedonism withstands evolutionary debunking arguments via what Street calls a "byproduct hypothesis." Since belief in pleasure's goodness is a byproduct of phenomenal introspection, which is selected for reliability, it's reliably caused even if other moral beliefs aren't. (ibid.)

Ultimately, he assumes, our hedonic starting points are not contaminated. This response to debunking thus relies on the same epistemic principle articulated by debunking itself, tackling it on its ground by picking out an epistemically reliable process that leads us to the intrinsic badness of suffering.

Alas, the strategy runs into problems of its own. One road to scepticism could simply be endorsing Nietzsche's view of consciousness as essentially epiphenomenal and deceptive (Riccardi 2015; 2018), but this would beg the question in Nietzsche's favour. Another way is to note that Nietzsche's view has found empirical support in the psychological literature and is echoed by a growing number of philosophical accounts of the limits of introspection. The fundamental thesis at odds with the introspective strategy is what Mattia Riccardi calls *Inner Opacity*, and one version of this objection is Eric Schwitzgebel's (2008) two-fold view: (i) that introspection is a multifarious composite of cognitive processes rather than a single unified faculty, typically failing to yield an informative judgment about the nature of our experiences; (ii) that whenever we do gain information, it is more often than not inaccurate or ambiguous, hence that introspection is generally unreliable. Schwitzgebel notes how difficult it is to know whether, say, joy has a single, distinctive, experiential character, and the same generalizes other affective states, and beyond. Nietzsche holds (i) and (ii) (e.g., D 35; GS 354; Riccardi 2015). But note that (ii) is sufficient to undermine the introspective awareness strategy. Schwitzgebel writes:

> Most people are poor introspectors of their own ongoing conscious experience. We fail not just in assessing the causes of our mental states or the processes underwriting them; and not just in our judgments about nonphenomenal mental states like traits, motives, and skills; and not only when we are distracted, or passionate, or inattentive, or self-deceived, or pathologically deluded, or when we're reflecting about minor matters, or about the past, or only for a moment, or where fine discrimination is required. We are both ignorant and prone to error. … even in favorable circumstances of careful reflection, with distressing regularity. (p. 247)

Schwitzgebel is sceptical that any unified account of the introspective phenomenology of, not just affective, but most mental states is forthcoming.[12] Could pain be an exception? Sinhababu suggests that Schwitzgebel "concede[s] that we can reliably introspect whether we are in serious pain." (p. 19) But this is a partial reading. Schwitzgebel writes about this "favorite example for optimists about introspection":

> [T]o use these cases only as one's inference base rigs the game. And the case of pain is not always as clear as sometimes supposed. There's confusion between mild pains and itches or tingles. There's the football player who sincerely denies he's hurt. There's the difficulty we sometimes feel in locating pains precisely or in describing their character. I see no reason to dismiss, out of hand, the possibility of genuine introspective error in these cases. (2008: 259-60)

Sinhababu also claims that Schwitzgebel's purported counterexamples rather show "that false beliefs about our experiences can be formed by reasoning about what we're likely to believe in a given situation, and not by phenomenal introspection." (p. 19) It is our reflective judgments that are prone

[12] Schwitzgebel differs from Carruthers (2010), who emphasizes the unreliability of introspective access to judgments and decisions, but not perceptual states. Still, even Carruthers' view would suffice to undermine access to our *beliefs* about the badness of suffering. Engelbert and Carruthers (2010: 241) also point out that "the cases discussed by Schwitzgebel do not provide much reason to think that our access to occurrent perceptual states is unreliable, even if we are poor at generalizing about our experiences or distinguishing causal from constitutive aspects of experience." But this only provides comfort for access to perceptual states. Again, our ability to introspect the value of suffering remains susceptible to scepticism in so far as it involves a *representation* of its badness (which it does on Carruthers' view). Further, Engelbert and Carruthers warn about the limited available data on introspection—in contrast to the abundant research documenting confabulation effects, namely, that "people will often interpret their own behaviour without awareness that they are doing so, and will, as a result, frequently make false self-attributions of mental states to themselves." (p. 251)

to error—confabulation or dumbfounding, one might say. Yet even if this were a correct reading, this would not guarantee reliable access, for we can't help but reflect on the value of our experiences to bring them to bear on our reasons. And so, we don't have reliable introspective access to (the evaluative judgment of) the badness of pain.

Thus, insulating judgments about the badness of suffering from debunking implies addressing such debates rather than presupposing the introspective route. Parfit's and Lazari-Radek and Singer's rationalist responses beg the question. Accordingly, if introspection fails to deliver on its promise, our attitudes to suffering remain susceptible to debunking. The worry is no longer simply that there is an evolutionary explanation of our evaluative attitudes; the full story now gives us positive reasons to deem the relevant processes unreliable. Indeed, they are not just non-truck-tracking, they are inherently falsifying.

Rebuttals of genealogical debunking thus fail. First, even if genealogy does not itself invalidate our beliefs, it undercuts their justification. Second, beliefs about the badness of suffering cannot avail themselves of the above three replies to debunking. At the very least, then, DBS does not provide a *better* explanation of our attitudes to suffering than naturalistic explanations. In the last section, I consider an additional argument, related to *Inner Opacity*, to tilt one's credence toward a Nietzschean account of our attitudes to suffering.

9. The Objection from Interpretation

Nietzsche conceives of moral judgments are "symptoms" or "sign-languages" of drives and affects (e.g., BGE 187; D34; 119; 542; TI "Problem": 2; "Skirmishes": 37; GM Preface: 2; Leiter 2013). That is, moral judgments are caused, and thus provide inferential evidence for, underlying drives. On Nietzsche's interpretive account, a claim like DBS is evidence of affective attitudes toward suffering. The diagnosis, a synchronic version of genealogy, does not falsify the belief, but it casts doubt on its independence from affects, hence on its epistemically defective source. For, on Nietzsche's view, we are disposed to have certain affective responses as a result of the way our drives (i.e., roughly, our subconscious urges) are organized. Drives are inherently interpretive and determine our orientation toward the environment. Paul Katsafanas (2013) comments:

> the affect influences the perceptual saliences, causing certain features to stand out and others to recede into the background. … In deliberation, the presentation of the facts – the selection of some features as salient and others as peripheral – is, at least in part, a function of the attitudes. (p. 741)
>
> Drives manifest themselves by coloring our view of the world … Nietzsche's idea is that the way in which one experiences the world is, in general, determined by one's drives in a way that one typically does not grasp (p. 743)

This explanatory framework accounts for our judgments because affects have valence: what we value as expressed by our judgments ultimately reflects what we value as a matter of motivational forces. Affects, as primarily noncognitive states, are not truth-apt.

However, commentators also argue that Nietzsche holds a two-level model of affective response. On one level, our responses involve phenomenal aspects *and* propositional attitudes. "Basic affects" are fully noncognitive, but we often display inclinations to and aversions from our basic affects, and these "*meta*-affects" may involve propositional attitudes. (Telech and Leiter 2017: 104)

If so, the worry is two-fold: basic affects are not truth-apt and meta-affects are susceptible to confabulation. Further, on Nietzsche's view, it is primarily custom that imposes a particular feeling on drives which, in themselves, are evaluatively neutral. (ibid.; cf. e.g., D 38) Accordingly, the underlying affective states of suffering are in principle distinct from the meta-affects held toward it, including our judgments about its value. Drives are, despite their valence, "morally undetermined" (ibid.). Only our "meta-affective stance (usually culturally shaped, and often involving beliefs) toward the basic affect" constitutes the moral valence of our sentiments (ibid.). But since we have no more reason to think that the way this stance was shaped was on track, scepticism persists.

This is Nietzsche's account. But a number of philosophers have argued that pain is only defeasibly bad. What does defeasibility mean in this context? We can take the cue from a particularist account of "defeasible generalizations" (Lance and Little 2004). The thought is this. We can at best generalize about pain. Its *default* status is to be bad, which explain ordinary attitudes, but generalizations are defeasible. Defeasible generalizations are thus powerful to explain both the variability and the pervasiveness of our attitudes to suffering.

Colin Klein (2014) argues specifically about pain that its default status does not impugn on the potential pleasantness of painful experiences. What motivates these arguments is the need to account for seemingly troubling cases like masochistic pleasures (broadly defined) (e.g., running a marathon, eating spicy food, a deep tissue massage). We can feel the painfulness of a pain as pleasant—a higher-order attitude toward a basically painful sensation that is on the edge of unbearability (recall the Nietzschean distinction between basic and meta-affects). On Carruthers' (2018) account, the valence of pain and other sufferings consists of a non-conceptual representation of the seeming badness of a sensation of pain. While Carruthers situates valence in the representation, not the sensation, his view also opens up a space for the valence of certain sufferings (i.e., the representation of their value) to vary. On the 'hedonic' theory, on the other hand, the valence of affective states is a qualitative property possessed by the very experience of those states. The 'hedonic' theory cannot account for dissociations between the sensations underlying affective states and the representation of these states (e.g., asceticism, masochism, pain-insensitivity).

Finally, Nietzsche's view of how our conscious moral sentiments are individuated finds echo in the provocative theory of emotions recently put forth by psychologist Lisa Feldman Barrett (2017), according to which emotions are the variable product of linguistic acculturation and inherently mediated by concepts superimposed on unindividuated affects. The individuation of different affective states under emotion *concepts* (predictive constructed prototypes; cf. ch. 5) does not pick out universal natural kinds with unique fingerprints. The emotions we experience, like sadness or *Schadenfreude*, are shaped by our cognition and our cultural upbringing. Further, Barrett explains why we have concepts to pick out certain states but not others. For the brain, "variation is the norm." (p. 282) When certain sensations are very intense or very frequent, we use such concepts to make sense of our sensory inputs, but what concepts we end up with is contingent. Thus, categorization, which relies on statistical regularities and language, shapes our conscious experience of the world.

Sounds familiar? Indeed, Nietzsche wrote: "[l]anguage and the prejudices upon which language is based are a manifold hindrance to us when we want to explain inner processes and drives: because of the fact, for example, that words really exist only for *superlative* degrees of those processes and drives." (D 115) "Extreme states" include "anger, hatred, love, pity, desire, knowledge, joy, pain". Our inner affective life is shaped linguistically and our concepts only crudely capture its fine-grained underlying nature. Our folk-psychological vocabulary, accordingly, is inevitably granular for we are only

introspectively aware of "extreme" or "superlative" states, and it is acquired through social interaction as members of a given community (Riccardi 2015). On Nietzsche's view, one is not "that which we appear to be in accordance with the states for which alone we have consciousness and words" (D 115).

One might still object that introspection does give us access to how we feel, at some level of generalization. Yet we just saw that there is no intrinsic, objective *axiology* to access in the first place. Our phenomenology of pain might (with caveats) track how we really feel, but judgments about its badness either are not reliably introspectable or are constructed. The challenge thus persists.

Conclusion

Many of our beliefs are the product of non-truth-tracking processes. While they may well be justified (say) pragmatically, this is an open question not settled by the intrinsic nature of suffering. Even if *Double Badness of Suffering* were justified, it would not vindicate *Realism about the Value of Suffering*. I laid out four debunking objections: peer disagreement, genealogical debunking, opacity, and interpretation, and argued that each response fails. Hence, RVS fails to securing the hope for normative convergence. The onus is now on the realist to explain why we should opt for less parsimonious explanations when a naturalistically sound account of our evaluative attitudes is available. Note that I did *not* argue that we are wrong to believe that suffering is bad, although I believe Nietzsche does. Instead, my argument sought to undercut RVS, a keystone of Parfit's argument for normative convergence. Not only does Nietzsche constitute a more serious threat than Parfit makes it seem; a new debunking challenge, drawing from Nietzsche, can be raised about RVS. Suffering matters, but it does not really matter in the realist's sense.

References

WORKS BY NIETZSCHE

The Anti-Christ, Ecce Homo, Twilight of the Idols and Other Writings, trans. J. Norman. Cambridge: Cambridge University Press, 2005

Beyond Good and Evil, edited by Rolf-Peter Horstmann and Judith Norman, trans. J. Norman. Cambridge: Cambridge University Press, 2002

Daybreak, edited by Maudemarie Clark and Brian Leiter, trans. R. J. Hollingdale. Cambridge: Cambridge University Press, 1997

The Gay Science, edited by Bernard Williams, trans. J. Nauckhoff and A. del Caro. Cambridge: Cambridge University Press, 2001

Kritische Studien-Ausgabe, edited by G. Colli and M. Montinari, 15 Vols. Munich: Deutscher Taschenbuch Verlag/de Gruyter, 1988

On the Genealogy of Morality, trans. M. Clark and A. J. Swensen. Indianapolis, IN: Hackett, 1998

OTHER WORKS CITED

Blackburn, S. (2011). Review of Derek Parfit, *On What Matters*, available at: http://www2.phil.cam.ac.uk/~swb24/reviews/Parfitfinal.htm

Bramble, B. (2017) Evolutionary arguments and our shared hatred of bread. *Journal of Ethics & Social Philosophy* 12(1):94-101

Carruthers, P. (2018) Valence and Value. *Philosophy & Phenomenological Research* 97:658-680

—— (2010). Introspection: Divided and Partly Eliminated. *Philosophy and Phenomenological Research* 80 (1):76-111

Clarke-Doane, J. (2012). Morality and Mathematics: The Evolutionary Challenge. *Ethics* 122(2): 313-40

Engelbert, M. and Carruthers, P. (2010). Introspection. *Wiley Interdisciplinary Review of Cognitive Science* 1(2):245-253

Gibbard, A. (2012) Five Girls on a Rock, *London Review of Books* 34(11)

Greene, J. (2013) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin Press

—— (2007). The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (eds.), *Moral Psychology, Volume 3*. Cambridge: MIT Press (pp. 35-79)

Haidt, J. (2000) The emotional dog and its rational tail: A social intuitionist approach. *Psychological Review* 108(4): 814-834

Hayward, M. K. (2018). Non-Naturalist Moral Realism and the Limits of Rational Reflection. *Australasian Journal of Philosophy* 96(4): 724-37

Huddleston, A. (2017) Nietzsche and the hope of normative convergence. In P. Singer (2017), pp. 169-94

Janaway, C. (2016) Attitudes to Suffering: Parfit and Nietzsche, *Inquiry* 20 (1-2):66-95

Joyce, R. (2016) *Essays in Moral Skepticism*. Oxford University Press

—— (2001) *The Myth of Morality*. Cambridge University Press

Kahane, G. (2016) Pain, experience, and well-being. In G. Fletcher, *The Routledge Handbook of Philosophy of Well-Being*. Routledge, pp. 209-220

—— (2014) Evolution and Impartiality. *Ethics* 124(2)

—— (2011) Evolutionary debunking arguments. *Noûs* 45(1):103-125

Katsafanas, P. (2013). Nietzsche's Philosophical Psychology. In K. Gemes and J. Richardson (eds.), *The Oxford Handbook of Nietzsche*. Oxford University Press, pp.727-55
Klein, C. (2014) The Penumbral Theory of Masochistic Pleasure. *Review of Philosophy and Psychology* 5(1):41-55
Klenk, M. (2018) Evolution and disagreement. *Journal of Ethics & Social Philosophy* 14(2)
Knobe, J. and Leiter, B. (2007) The Case for Nietzschean Moral Psychology. In B. Leiter and N. Sinhababu (eds.), *Nietzsche and Morality*. Oxford University Press, pp. 83-109
Lance, M. and Little, M. (2004) Defeasibility and normative grasp of context. *Erkenntnis* 61(2-3):435-455
Lazari-Radek, K. (de) and Singer, P. (2017) Parfit on objectivity and "The profoundest problem of ethics". In P. Singer (2017), pp. 279-96
—— (2013) *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford University Press
Leiter, B. (2014) Moral skepticism and moral disagreement in Nietzsche. *Oxford Studies in Metaethics: Volume 9*, pp. 126-151
—— (2013) Moralities are a sign-language of the affects. *Social Philosophy & Policy* 30: 237-258
Lin, E. (forthcoming) Attitudinal and phenomenological theories of pleasure, *Philosophy & Phenomenological Research*. DOI: 10.1111 / phpr.12558
Mackie, J. L. (1977) *Ethics: Inventing Right and Wrong*. Penguin Books
Nagel, T. (1986) *The View from Nowhere*. Oxford: Oxford University Press
Nichols, S. (2014) Process debunking and ethics. *Ethics* 124(4):727-74
—— (2004). *Sentimental Rules. On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press
Parfit, D. (1984) *Reasons and Persons*. Oxford University Press, 1984
—— (2011) *On What Matters,* volumes I and II. Oxford University Press
—— (2017) *On What Matters,* volume III. Oxford University Press
Riccardi, Mattia (2018). Nietzsche on the Superficiality of Consciousness. In Manuel Dries (ed.), *Nietzsche on consciousness and the embodied mind*. De Gruyter, pp. 93-112
—— (2015) Inner Opacity. Nietzsche on Introspection and Agency. *Inquiry* 58 (3):221-243
Schwitzgebel, E. (2008) The unreliability of naive introspection. *Philosophical Review* 117(2):245-273
Singer, P. (ed.) (2017) *Does Anything Really Matter? Essays on Parfit on Objectivity*. Oxford University Press
—— (2005) Ethics and intuitions. *Journal of Ethics* 9: 331-52
Sinhababu, N. (manuscript) The epistemic argument for hedonism
Skarsaune, K. O. (2011). Darwin and moral realism: Survival of the iffiest. *Philosophical Studies* 152 (2): 229-243
Sobel, D. (2011) Parfit's Case Against Subjectivism. *Oxford Studies in Metaethics: Volume 6*, pp. 53-78
Street, S. (2008) Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About. *Philosophical Issues* 18: 207-228
—— (2006) A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127 (1):109-66
Telech, D. and Leiter, B. (2016) Nietzsche and moral psychology. In J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy*