**Accepted manuscript. Forthcoming in *Inquiry: An Interdisciplinary Journal of Philosophy*. Please cite and refer to the published version when available at https://www.tandfonline.com/toc/sinq20/current**

# Strangers to ourselves: A Nietzschean challenge to the badness of suffering

Nicolas Delon[1]

ABSTRACT. Is suffering really bad? The late Derek Parfit argued that we all have reasons to want to avoid future agony and that suffering is in itself bad both for the one who suffers and impersonally. Nietzsche denied that suffering was intrinsically bad and that its value could even be impersonal. This paper has two aims. It argues against what I call 'Realism about the Value of Suffering' by drawing from a broadly Nietzschean debunking of our evaluative attitudes, showing that a recently influential response to the debunking challenge (the appeal to phenomenal introspection) fails. It also argues that a Nietzschean approach is well suited to support the challenge and is bolstered by the empirical literature. As strangers to ourselves, we cannot know whether suffering is really intrinsically bad for us.

## 1. Introduction

Both Nietzsche and Parfit spent much of their lives concerned with suffering, the meaning of life, and whether it has all been worth it.[2] My goal in this paper is two-fold: to motivate the view that the badness of suffering does not turn on its intrinsic nature and to marshal a Nietzschean challenge to the idea that the badness of suffering is both intrinsic and mind-independent. I focus on the late Parfit's confrontation with Nietzsche[3], and the question of how, if ever, we can know that suffering is bad in light of Nietzsche's enigmatical remark that "we remain of necessity strangers to ourselves" (GM, Preface, 1). My main target will be introspection, and what

[1] Nicolas Delon, New College of Florida. Contact: ndelon@ncf.edu. Thanks to Gwen Bradford, Carter Delegal, François Jaquet, Andrew Lee, Brian Leiter, Duncan Purves, Yann Schmitt, Neil Sinhababu, an audience at the 65th meeting of the Florida Philosophical Association, and several anonymous referees. The detailed comments of a referee for *Inquiry* contributed a great deal to improving this article, which I dedicate to the late Clément Rosset.

[2] See Bernard Williams' introduction to *The Gay Science*, p. xiv. Williams also remarks on Nietzsche's "hypersensitivity to suffering".

[3] See Huddleston (2017); Janaway (2016) for previous discussions.

Nietzsche and cognitive science tell us about its reliability. The discussion takes Nietzsche as bona fide interlocutor and sheds light on his broader project of revaluation of our values. I depart from extant treatments of his views on suffering and self-knowledge, respectively, but also draw a bridge across them. Whereas commentators (understandably) focus on his metaethical positions when dealing with the value of suffering, I bring Nietzsche's account of consciousness and self-knowledge to bear on this very question. This, in turn, sheds new light on his reasons to reject our common attitudes to suffering.[4]

Start by considering a platitude and mainstream view: pain is unconditionally, intrinsically bad (Goldstein 1989).[5] Pain is also widely considered a source of reasons: from its badness follows that we ought to (and it would be morally wrong not to) prevent, reduce, relieve, or not cause it, other things being equal. Indeed, pain seems like a paradigm case of intrinsic badness. It's not just pains, headaches and stubbed toes, mild, stabbing, throbbing, chronic, or acute pains, that hurt. Psychological suffering, emotions, moods, anxiety, depression, also hurt. And we mind not only our suffering, but others' suffering. On the other hand, we are familiar with instrumentally good suffering (say, when getting a flu shot or training for a marathon), and some that we seem to enjoy for themselves (as in a deep tissue massage, a very spicy dish, the burn of a hard workout, or the pangs of grief). Yet philosophers haven't seemed bothered by such counterexamples.[6] The intrinsic badness of pain or suffering seems to be among philosophy's safest truisms.

[4] Nietzsche uses "suffering" (*Leiden*), "pain" (*Schmerz*), and "displeasure" (*Unlust*/*Unmut*) interchangeably. For clearer distinctions, see Reginster (2006: 113-114; 234-235).
[5] I am concerned with suffering, not just pain. Pains are nociceptive sensations with a certain phenomenal character: unpleasant, hurtful. Sufferings are strongly valenced affective states (mundane unpleasant experiences like folding the laundry do not rise to the level of suffering – thanks to an anonymous referee for making this point). I will assume that pain, as involving both sensory and affective components, is one kind of suffering.
[6] A few exceptions: Bradford (2020), Brady (2018), Carel and Kidd (2020), Klein (2014), Kraut (2007: 151-157), Lance and Little (2004). Also see Leknes and Bastian (2014) for empirical discussion.

Here is a metanormative thesis about suffering:

> ***Realism about the Value of Suffering*** **(RVS)** The value of suffering is independent of our attitudes toward it.

RVS tells us there is a correct attitude to an occurrence of suffering (indeed, to all tokens of a suffering-type). This judgment is independent of what view of pain one holds, whether its badness consists in how it *feels* or what *attitudes* we take to its properties. Value in RVS is attitude-independent in a strong sense. Of course, the value of a token of pain is constituted by the subject's mental states, at some level. RVS says that *evaluative* attitudes such as approval or disapproval, liking or disliking, and desires do not determine whether pain is bad for its subject. I will argue that RVS, and related claims that presuppose it, are harder to establish than it seems. I proceed as follows. §2 introduces the Parfit-Nietzsche debate and lays out the stakes by spelling out the relation between suffering and reasons. §3 details Nietzsche's views on suffering. §4 motivates the debunking challenge. §5 describes a defense against the challenge, the introspection strategy. §6 argues from the unreliability of introspection that the strategy fails. §7 ties together converging lines of argument from Nietzsche and cognitive science to debunk the apparent truism at hand.

## 2. Parfit on disagreement about suffering

In the second volume of *On What Matters* (OWM), Parfit dedicates more than a full chapter to Nietzsche, after discussing moral disagreement and convergence. He describes him as "the most influential and admired philosopher of the last two centuries" (OWM II 570; cf. 1984: 176). Parfit argues for a substantive view (as opposed to the metanormative RVS):

> ***The Double Badness of Suffering*** **(DBS)** All suffering is in itself both bad for the sufferer and impersonally bad. (II 569)

RVS is the metaethical backbone of DBS. They could come apart, but the badness of suffering is the clearest case of what Parfit calls "irreducibly normative truths" throughout *On What Matters*. Our reasons to want pain or agony not to occur are primitive; they do not depend on our attitudes or other facts. The very idea of badness in itself is of course suspect to Nietzsche, since any suffering assumes a sufferer, and different things are good for different people. What Parfit means is that the overall badness of suffering is not only a function of its prudential badness for the sufferer; it is also bad impersonally (or as a state of affairs) that the sufferer suffers — "from the point of view of the universe" as Sidgwick puts it. Moreover, for Parfit, all sufferings matter equally. That is, all sufferings of comparable magnitude are equally bad regardless of who is suffering. This view is best described by:

> ***Impartiality*** Suffering is a source of impartial, or "person-neutral," reasons: i.e. no reference to any particular person attaches to the explanation of why one ought to reduce it. We *all* have reasons to regret, prevent or relieve the suffering of *anyone* (OWM I 138; cf. Nagel 1986: 161).

Clearly, Nietzsche rejects *Impartiality*. First, he denies that that value is mind-independent and that it is invariant (e.g. BGE 2; GS 299, 301; TI, Skirmishes, 19, 49). The two conjuncts are important for the value of suffering could be nearly always negative even if it were mind dependent. Nietzsche's anti-realism commits him to mind-dependence; his view of *how* value is created (roughly, projected) leads to denying invariance. Nietzsche can thus claim that suffering *can* be valuable. As part of his project of revaluation of our values, Nietzsche is committed to a

"*radical* revaluation of the role and significance of suffering in human existence" (Reginster 2006: 44).[7]

Parfit's main argument presupposes reliable access to DBS – a key example of irreducible normative truth. I will argue that the presupposition is unwarranted. As Huddleston (2017) aptly says, Parfit is throwing stones from a glass house against Nietzsche. What stones, exactly? Let me now reconstruct Nietzsche's position.

## 3. Nietzsche on the value suffering

For Nietzsche, suffering encompasses a great variety and degrees of physical and psychic states (physical pains, guilt, disgust, sadness, sorrow, shame, loneliness, disappointment, dissatisfaction, illness; imposed by bad luck, cruelty, asceticism or punishment). As noted, his central project has as its "ultimate object of revaluation … the role and significance of suffering", in part as a response to nihilism (Reginster 2006: 185). Nietzsche's doctrine of *amor fati* can be seen as central this revaluation in that justifying existence may necessitate the affirmation of suffering: to love one's life as a whole, "saying yes of life even in its strangest and harshest problems" (TI, What I owe to the Ancients, 5). His "formula for greatness" states: "you do not want anything to be different, not forwards, not backwards, not for all eternity" (EH, Why I am

[7] If Nietzsche's project is coherent, it threatens Parfit's hope that philosophers would in nearly ideal conditions converge on fundamental normative truths, such as DBS, which is why "we cannot ignore Nietzsche" (OWM II 26). Parfit makes a sweeping empirical prediction:

> *The Convergence Claim*: 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)

That includes DBS and *Impartiality*. If Nietzsche does not meet the conditions, as Parfit argues, because his claims reflect distorting influences (i.e. he lacked a proper concept of reasons, had a taste for hyperbole, was engaged in motivated reasoning, was succumbing to dementia…), then he doesn't really threaten *Convergence*. For Parfit, Nietzsche's exaltations of suffering "conflict deeply with what most of us believe." He only "tried to believe that suffering is not bad" (OWM II 571). Textual support for this interpretation is questionable (Huddleston 2017; Janaway 2016).

so clever, 10[8]). The value of suffering, on this reading, is neither merely instrumental nor strictly intrinsic or final. It is not suffering *per se* that one must affirm, nor is it suffering that *causes* otherwise valuable things; it is a whole sequence of which suffering is a necessary constitutive ingredient (Janaway 2016; Mollison 2018).

To appreciate the coherence of Nietzsche's position, let us unpack its moving parts. First, Nietzsche construes (at least one major type of) suffering as the feeling of resistance. Since life values overcoming resistance, an expression and enhancement of power, life values suffering as a necessary component of what is ultimately valued (Reginster 2006: 177). Put differently, agency, largely subconsciously, values obstacles to the will beyond the mere satisfaction of desires. This much, I assume, is highly plausible: difficulty, effort, and resistance are integral parts of what gives our actions value – they are at least part of what the will to power aims at (for a nuanced criticism, see Clark 2012). Reginster's interpretation doesn't require controversial adherence to a metaphysical doctrine of the will to power as the nature of life or reality (e.g. BGE 36). He endorses to a large extent Clark's (2000) account that Nietzsche's valuations are ways of looking at the world from the viewpoint of psychological drive of the will to power. Further, like Clark, Reginster underscores the will to power as a *second-order* drive, not all other drives being directed at power. In other words, human agency is largely motivated by a strive to maximize feelings of power (e.g. GM III 7).[9]

Secondly, Nietzsche repudiates the Schopenhauerian notion that happiness is pleasure is absence of resistance (Katsafanas 2015b).[10] Hedonism about well-being rests on a faulty psychological

[8] Relatedly, eternal recurrence provides a counter-ideal and a test of one's ability to endure the horrors of existence (EH Z 1; GS 341; Leiter 2015: 230)
[9] For more discussions of the doctrine, see e.g. Katsafanas (2013), Schacht (1983), and Richardson (2004).
[10] Schopenhauer (1969 I: 363): "all suffering is simply nothing but unfulfilled and thwarted willing".

theory. Even Nietzsche's preferred positive affect, joy or cheerfulness (*Fröhlichkeit*), is sometimes said to be a mere side effect of power increase.[11] As I will argue, phenomenal opacity poses a further challenge to hedonism. In contrast to utilitarian and Epicurean hedonism Nietzsche's "new happiness" rests on higher values and "great pain" (GS, Preface, 3).

> Hedonism, pessimism, utilitarianism, eudaemonism: these are all ways of thinking that measure the value of things according to *pleasure* and *pain,* which is to say according to incidental states and trivialities. … You want, if possible … *to abolish suffering.* And us? – it looks as though *we* would prefer it to be heightened and made even worse than it has ever been! Well-being as you understand it – that is no goal (BGE 225)[12]

Nietzsche's claim is not that well-being is not normative, which would trivially falsify DBS and *Impartiality*. Instead, he proposes a "higher values" conception of well-being or happiness. This requires a positive account of the value of suffering, the third moving part of Nietzsche's position.

We just saw that the feeling of resistance plays a role in the will to power's valuations. But the supervenience relation between value and resistance is complex. The value of a token of suffering depends on its being token of overcoming as well as on its role within a particular sequence. One of Nietzsche's metaphors for life, if not the world, was that of a work of art (e.g. BT 5, GS 290; cf. Nehamas 1985 for a defense and Leiter 1992 for criticism). More relevant for my purposes,

[11] It is plain that for Nietzsche pleasure is not our main or deepest aim. "Man does *not* seek pleasure and does *not* avoid displeasure. … Pleasure and displeasure are mere consequences, mere accompanying phenomena—what man wants … is an increase of power." (KSA 13:14[174]) Kaufman (1974: 262) argues that pleasure and pain are epiphenomenal.

[12] Nietzsche mocks utilitarians for "striving for English happiness, I mean comfort and fashion" under the guise of virtue (BGE 228). As Janaway (2016: 81) notes, "when Parfit begins his whole project with the statement that 'on any plausible theory [of well-being] hedonism covers at least a large part of the truth' [OWM I 40], real disagreement is already brewing."

much of what Nietzsche writes about the value of suffering bears on character development, personal growth and heroic accomplishments. In light of this, we can best appreciate suffering's value as part of a "compelling narrative of adversity and achievement" (Huddleston 2017: 179). Relatedly, Janaway writes: "If a course of events … instantiates as a whole what we can call *growth-through-suffering*, it may contribute to a recognized form of well-being. Such a course of events can both be *good as a whole*" (2016: 83). But Nietzsche himself puts it best: "Sometimes the value of a thing is not what you get with it but what you pay for it" (TW, Skirmishes, 38).[13]

Taking a different tack, Leiter (2018) argues that, in response to the recognition that "truth is terrible" (including that death and suffering are pervasive and inevitable), and to Schopenhauer's question "Why continue living at all?", Nietzsche answers that life can only be justified aesthetically, where "justification" means "restoring an affective attachment to life". Indeed, Nietzsche praised artists for facing even glorifying what is "ugly, harsh, questionable in life"; "the *heroic* man praises his existence through tragedy" (TI, Skirmishes, 24; cf. §38) "What makes one heroic?—Going out to meet at the same time one's highest suffering and one's highest hope" (GS 268). However, Leiter's explanation of the efficient causes of this restoration is controversial. He argues that Nietzsche assimilates aesthetic pleasure to a kind of sublimated sexual pleasure, and that powerful affects can neutralize pain. Leiter is right that nihilism, and its best treatment, are primarily affective. But Nietzsche's view that suffering can be desirable remains on the table. Indeed, the narcotic power of aesthetic pleasure is only valuable if it does not numb the affirmation of the harshness of existence. This is borne out by the following observations. The ascetic ideal sustains destructive suffering while numbing its adherents (GM

[13] See Reginster (2006: 103-47, 177). Textual support includes BGE 230, 259; GS 13, 338; A 2; KSA 11 [111] 13; 14 [173] 13; 14 [174] 13; 11 [75] 13; 9 [151] 12.

III; cf. GS 56 on young Europeans' "yearning to suffer something in order to make their suffering a reason for action"). Romantics suffer from an "*impoverishment of life* and seek quiet … redemption from themselves through art and insight, or else intoxication, paroxysm, numbness, madness", whereas others suffer from a "*superabundance of life* [and] want Dionysian art as well as a tragic outlook and insight into life" (GS 370). It is difficult to make sense of these important claims without assuming that suffering can be more than just instrumentally valuable and ultimately worth neutralizing. I thus agree with Reginster that, "[f]rom the standpoint of the ethics of power, suffering cannot be coherently deplored". (Reginster 2006: 233)

Tying these threads together, I submit that the following formula best encapsulates Nietzsche's position:

> ***Narrative value*** An episode of suffering is valuable when, and insofar as, it can be incorporated into an aesthetically compelling sequence of achievement that required overcoming resistance.

A corollary of *Narrative value* is that the value of suffering cannot be determined prior to the development of the relevant sequence(s). This is enough to reject the notion that its value is intrinsic. But now we know that Nietzsche suggests that suffering can be valued *non-instrumentally*, not just extrinsically. What does this mean? Reginster (2006: 15) writes that the revaluation of life-negating values "must show that those aspects of human life condemned by the nihilist … are not only bearable, but also desirable, and not only derivatively, but for their own sake." But does he literally mean that suffering must have *final* value?

John Richardson argues that Nietzsche's view involved "making *every* case of suffering good … even the physical agony of those quite unable to overcome it and grow through it" (2015: 103).

This goes further than we have conceded so far, and Richardson's reading is heavily indebted to the *Nachlass*. Nietzsche is also adamant that not everyone is apt to turn their suffering into something good. While his view sometimes does sound that strong, there is a more plausible reading (Delon 2019; Janaway 2016; Mollison 2018; Reginster 2006): the value of an experience is non-invariant and can be conferred meaning within a coherent whole – akin to how artists can make ugly things beautiful (GS 299) or how Achilles' noble character implies traits that would be despicable in isolation (GS 79) (Hassan, forthcoming). More precisely, following Hassan's amendments to Reginster's ambiguous position, we can make the following distinction. Suffering is not an invariantly regrettable *enabler* of an overall positively valuable whole. Rather, it is a *contributor* whose value varies across contexts and cannot be coherently deplored when part of a valuable organic unity.[14] For "everything is redeemed and affirmed in the whole." (TI, Skirmishes, 49) Importantly, on each of these possible interpretations, suffering can be valued non-instrumentally. Thus, *Narrative value* seems well supported. Nietzsche did not hold, *pace* Parfit, the weaker view that suffering could only be instrumentally good (also *pace* Leiter 2015: 103-8).

Parfit could reply that aesthetic value does not affect the *moral* badness of suffering, and second, that suffering only contributes to compelling narratives *because* it is bad (OWM III 309; Kahane 2016: 217). But since Nietzsche operates an explicit rapprochement, if not conflation, between moral and aesthetic spheres (e.g. Came 2014; Hassan, forthcoming), the first point is moot. The second point rests on a confusion. Suffering's contribution does not hinge on its badness, which

[14] Hassan borrows the distinction from Dancy (2003).

is underdetermined by its intrinsic properties, but on its tendency, as feeling of resistance, to stimulate the feelings of power.

Another objection could be that most of us do not, in fact, consciously value resistance. Indeed, our decadent culture is characterized by "*the religion of snug cosiness*" (GS 338); we no longer accept pain. GS 48 deplores modern culture's "general inexperience" with distress and pain (cf. BGE 206). If we value suffering, it is subconsciously. Nietzsche is happy to concede this observation. What matters is what role suffering can play, and how we can learn to embrace it. Nietzsche wants us to value it more *deliberately* and promotes the cultural achievements that had suffering at their core. At a deeper level, humanity does value suffering's capacity to create, transform, sublimate (BGE 225). The observation speaks to the problem that Nietzsche diagnoses. He explains the rejection of suffering as a remnant of religiosity and a sign of a new "religion of compassion" (GS 338). Modern European culture promotes a "tyranny of timidity" (D 174), "a deadly hatred against suffering in general", "a new Buddhism" (BGE 202; cf. KSA 12, 9 [82] and GM, Preface, 5). Europeans wish "the universal, green pasture happiness of the herd, with security, safety, contentment, and an easier life for all … 'equal rights' and 'sympathy for all that suffers'" (BGE 44; cf. BGE 270; D 174; GS 377). The ascetic ideal, more broadly nihilism, interprets our (previously meaningless) suffering as constituting "an objection to life" (EH, Z, 1), "a defect of existence" (GS 338).

Nietzsche's case for affirmation also rests on what is lost in the reduction of suffering:

> The plant 'man' has grown the strongest … under conditions that are quite the reverse (BGE 44).

> The discipline of suffering, of *great* suffering—do you not know that only this discipline has created all enhancements of man so far? (BGE, 225)

Alongside observations of decadence and nihilism lies a plea for higher values. Suffering is not just a necessary, if regrettable, means to the likes of Goethe, Beethoven, or Napoleon, or the construction of Venice. It is *constitutive* of these achievements. *The Gay Science* is probably the richest trove on the revaluation of suffering. For instance:

> But what if pleasure and displeasure are so intertwined that whoever *wants* as much as possible of one *must* also have as much as possible of the other …? (GS 12)

There, Nietzsche envisions "*as much displeasure as possible* as the price for the growth of a bounty of refined pleasures and joys that hitherto have seldom been tested" (ibid.). "There is as much wisdom in pain as in pleasure: like pleasure, pain is one of the prime species-preserving forces … that it hurts is no argument against it" (GS 318).

On the other hand, pity (*Mitleid*) hampers these positive contributions. "Nietzsche objects to pity. No fact about his critique of morality is so widely known", wrote Nussbaum (1994: 139). Nietzsche's objection is complex. What matters for present purposes is what the critique of compassion or pity says about his reasons for affirming suffering. Most sufferers are not worthy of beneficence; nor is compassion good for most of them, let alone for those who experience it (BGE 293; Z II, On the Pitying). Relevant to *Narrative value*, compassion "*strips* of the suffering of what is truly personal", while there is a "personal necessity of misfortune". When properly incorporated suffering can contribute to the "whole inner sequence and interconnection that spells misfortune for *me* or for *you*!", from which an individual draws strength and self-understanding (GS 338; cf. Janaway 2016; Leknes and Bastian 2017). Suffering is thus essential to individuality

as well as knowledge. “Profound suffering makes you noble; it separates”; “anyone who has suffered deeply” acquires “a trembling certainty that his sufferings have given him a *greater knowledge* than the cleverest and wisest can have” (BGE 270; cf. GS 55 on the singularity of the noble).[15]

But note that such statements underscore the scope of Nietzsche’s view: you cannot infer the value of experiencing *x* for everyone from *your* experience of *x,* because it is inevitably indexed to your personal situation. This cuts both ways: suffering need not be bad; but its being good for some does not imply it can be good for everyone (BGE 228). “The higher nature’s taste is for exceptions, for things that leave most people cold and seem to lack sweetness; the higher nature has a singular value standard.” (GS 3) It may be that only outstanding individuals are prone to find joy and excellence in suffering, and culture should promote such individuals (perhaps even at the cost of great in the many who can’t turn it into something good; cf. BGE 257, 258; Huddleston 2014). Regardless, Nietzsche suggests a number of grounds on which suffering can be made non-instrumentally good, whether it be at the level of the individual level or society.

To sum up, episodes of suffering have a *contributive* value that depends on their role in a meaningful narrative.[16] Key to this possibility is a point at odds with RVS, namely that the

---

[15] The topic of self-discovery through suffering is broached many times in *The Gay Science* (e.g. Preface, 12, 290; 326, 338, 347, 370; cf. D 174 and GM I 10, III 1).

[16] Suffering requires also “meaning” in distinct sense.

> [Man] suffered from the problem of his meaning. … that the answer was missing to the scream of his question: ‘*to what end* suffering?’ … The meaninglessness of suffering, not the suffering itself, was the curse that thus far lay stretched out over humanity—*and the ascetic ideal offered it a meaning*! (GM III 28; cf. II 7; cf. D 32 and GS 56)

See e.g. Janaway (2007: ch. 13); Leiter (2015: ch. 8); McPherson (2016); Reginster (2006); Williams (2006). GM III does give fodder to the idea that the value of suffering is malleable and a function of our interpretation. It also suggests that our belief in the badness of suffering is skewed by focusing on *meaningless* suffering, which Nietzsche deplores, if only because it leads to nihilism. Saying “yes” to everything cannot mean resignation in the face of meaningless suffering. Unlike Schopenhauer, however, Nietzsche didn’t see their suffering as an objection to this world (Reginster 2006). Still Nietzsche was keenly aware of the problem. In *Schopenhauer as Educator*, he wrote: “More profoundly feeling people have at all times felt sympathy for the animals because they suffer from life and yet

attitudes we take toward suffering shape its contribution. This applies independently of whether, or in what sense, it really is *up to us* what attitudes we hold. In fact, *Narrative value* is compatible with the idea that we are essentially unfree to write our own stories, that our deep character or "type" dictates our attitudes, but that's a topic for another day (cf. Knobe and Leiter 2007). Since for Nietzsche all value is conferred upon things (GS 301), it is in some sense trivially true for him that the value of suffering is our own making. Let's be more precise. Second-order attitudes can affirm suffering, *including our first-order desires not to suffer*. Indeed, the will to power "has the structure of a *second-order desire* … for the overcoming of resistance in the pursuit of some determinate first-order desire." (Reginster 2006: 132; cf. Clark 2000; 2012) Again, it may not be up to us what second-order desires we develop, but within the economy of power they determine how much of it we can express and increase.

I have canvassed two contrasted positions on the value of suffering. Parfit assumes we can assume claims like DBS, but what reasons do we have, if we want to take Nietzsche seriously, to actually hold (or doubt) such claims? The remainder of this article mounts a debunking challenge to DBS and the philosophical commonsense about suffering.

## 4. Debunking

Let me start by describing the relation between debunking arguments and our moral beliefs in general. Debunking arguments have causal and epistemic premises. Evolutionary explanations of our beliefs (causal premise) can undermine objectivism or realism about values by casting our beliefs as systematically prone to error, modally contingent or better explained by non-truth-

do not possess the power to turn the thorn of suffering against itself and to understand their existence metaphysically; one is, indeed, profoundly indignant at the sight of senseless suffering. (UM III 5) Parfit (OWM II 592; III 311) misinterprets the passage as suggesting that Nietzsche attached equal importance to the suffering of animals.

tracking processes (epistemic premise): "Off-track influences on the belief that *p* undermine our justifications for believing that *p*." (Kahane 2011: 106) We need an account of reliability.

A process is reliable (i.e. truth-tracking) if and insofar as it generally tends to produce true beliefs (in the target domain) that we can justify. A process that produces significantly more false than true beliefs is unreliable, and when we know that our beliefs are produced by such processes, we have reason to think they may be false. When it is reliable, the fact that we have beliefs in the relevant domain gives us appropriate reasons to think they are true (Nichols 2014; Tersman 2017: 758). Motivated reasoning, confabulation, situational influences, subconscious affects, among others, can shed doubt on our justifications. Call them *undercutters*.

Debunking operates along two dimensions:

> ***Synchronic*** Insofar as our evaluative beliefs depend on "epistemically defective emotional and motivational processes" (Nichols 2014) their justification is undercut.
>
> ***Diachronic*** Insofar as our moral judgments and norms can be traced to their cultural and environmental origins (say, among our hominid ancestors), their justification is undercut (e.g. Joyce 2001; Street 2006).

The explanation of how we moralize our responses to violations of prosocial norms is only contingently related to what we take to be the moral truth. The evolution of our affects has equal if not greater explanatory power than assuming we can detect objective truths. Nichols (2004), for instance, offers a compelling story of how "harm norms" have evolved to reflect widespread changes in sensibilities, themselves explained by sociocultural factors. We might have ended up with radically different moral beliefs, and contingency runs afoul of a deep commitment of

ordinary moral discourse to categorical moral reasons (Joyce 2001, ch. 2). If we the evolutionary processes that gave rise to our moral beliefs are unreliable because they are not truth-tracking, then even minor variations along the evolutionary path would have led to mostly false moral beliefs (Street 2006). Plausibly, debunking is stronger when it features both diachronic and synchronic undercutters, as Nietzschean moral psychology will illustrate. I will argue that analogous moves are available to debunk DBS and *Impartiality*. Beforehand, a caveat is in order.

Debunking does not establish the falsity of our beliefs; rather, it undercuts our justifications for holding them absent countervailing evidence. We may have good evolutionary explanations of the norm against murder; this doesn't mean we *cannot* justify the norm. Expressivists, constructivists and sentimentalists all propose tools not just to explain but also *justify* our moral norms. The debunking arguments we are interested in are epistemological (Joyce 2016: ch. 8; Vavova 1015: 105), namely, they target our justifications for holding certain beliefs.

The contingent history of "harm-norms" regarding, say, animal cruelty or punishment could explain the rise of hedonism and utilitarianism (Nichols 2004: ch. 6-7). Is there a debunking explanation of Parfitian attitudes? Nietzsche's genealogy traces our attitudes to their physiological causes, explaining morality as a "sign-language of the affects" (BGE 187). DBS could be a symptom of hypersensitivity to suffering explained by historical and sociological facts. For Nietzsche, the morality of compassion is "just another expression of … physiological overexcitability" (TI, Skirmishes, 37). In fact,

> Nietzsche might just as well claim … 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' make them fetishize the phenomenal character of suffering (Huddleston 2017: 180)

So why assume that Parfit is *not* susceptible to distorting influences? Parfit follows a long moral tradition but could be oblivious to his own sensibilities. He could be susceptible to diachronic and synchronic undercutters. I will argue that a critical route to justifying RVS and DBS does not withstand debunking.

## 5. The introspection strategy

### a. Debunking and reason

An initially plausible defense for realists and objectivists lies in an appeal to our cognitive faculties. It goes like this: if there is one set of attitudes that cannot be readily debunked, these are our attitudes to pain. To see why the reliability of cognitive faculties is key, consider Kahane's glasshouse objection to the use of evolutionary debunking arguments to vindicate utilitarianism and *Impartiality* (e.g., Singer 2005):

> 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. … [I]f anything would survive [the doxastic purge], 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. … worrisomely, the view that emerges in outline is more Nietzsche than Singer. (Kahane 2011: 120)[17]

After all, there is "no mystery whatsoever" in the emergence of our attitudes to painful sensations "associated with bodily conditions such as [cuts, burns, bruises, broken bones]": they enhanced our survival and reproductive fitness (Street 2006: 150). In contrast, the belief that pain is *good* and pleasure *bad* cannot be explained by similar mechanisms. Street argues that, since our "basic evaluative tendencies" were shaped by evolutionary forces unrelated to the moral truth, most moral judgments may be off-track, and so the realist confronts a dilemma: either an inexplicable coincidence took place (evolutionary pressures shaped our evaluative attitudes just so they reached the truth) or our faculties can reliably track those truths (a naturalistically implausible take). But since the truth of our beliefs need not figure into the explanation of why we hold them, for we were not selected to have true moral beliefs, we would have believed that pain is bad whether or not it is bad in the realist's sense. On an anti-realist account like Street's, pain is bad simply because of it is the object of our negative attitudes. The anti-realist can hold that pain is really bad in a sense; they just also hold that it could have been otherwise.

Parfit views many moral beliefs as immune to debunking because they are discovered by reason, like logical and mathematical truths. He and Lazari-Radek and Singer (henceforth LRS) (2014; 2017) respond to Street's dilemma that rational reflection exerted a countervailing, reliable pressure on the formation of our moral beliefs, countering the epistemic premise of debunking. LRS (2017) argue that a Principle of Universal Benevolence, akin to *Impartiality*, survives debunking for similar reasons. Likewise, for Parfit, the belief that everyone's well-being matters

[17] Cf. Kahane (2014: 330-2); (2016: 217-8). Nonetheless, Kahane (2016) argues that pain *is* intrinsically bad and does so by appealing to the inevitability of how it feels (phenomenal introspection, that is).

equally, unlike the belief that incest is wrong, could not have been reproductively advantageous; (OWM III 340).[18]

Does the strategy hold water? Street denies that reason could put us on track. For the defense assumes that rational reflection provides uncontaminated tools to separate the wheat from the chaff. 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." (Street 2006: 124) In other words, "rational reflection cannot turn muck into gold." (Hayward 2018: 726)[19]

LRS press on. We have reasons to trust, not our starting points, but our faculty of reasoning. Its origins do not impugn the content of the beliefs it produces. Our ability to discern objective moral truths is a by-product of an adaptive package of rational capacities that could not be "economically divided" (2017: 288). What to do of Street's claim that 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" (2006: 124)? LRS claim that we *can* sort the wheat from the chaff. Granted, we hold some of the beliefs we do because they were adaptive for our ancestors, but some beliefs are not amenable to such explanations, such as Sidgwick's Principle of Universal Benevolence. Such beliefs belie evolutionary debunking because they could *not* have been adaptive for our ancestors, yet we can recognize their truth. However, many people do reject this allegedly "self-evident" principle, and as we saw, there might still be factors of cultural evolution and history that explain why such a principle appears to be true. Why, then, presume that

[18] Jaquet (2018) argues that we acquired our prudential beliefs (about well-being) through our rational capacities rather than evolution.
[19] Skarsaune (2011) offers a conditional argument that, *if* pain is bad, then we can offer a causal story of why we reliably came to form justified beliefs about its badness. He offers no argument that pain *is* bad.

reflection would lead us to hedonism about well-being rather than, say, Nietzschean perfectionism[20] (Kahane 2011)? Parfit and LRS must eventually presuppose a failsafe process: introspection, a source of uncontaminated content from which rational reflection can proceed. That is, the defense only works if we presuppose both the existence and the accessibility of some basic normative fact. We can learn through phenomenal introspection, the argument goes, that suffering is intrinsically bad: "Our beliefs in the goodness of pleasure and the badness of pain themselves … are grounded in a different way from our other moral beliefs." Our "direct acquaintance" with these states of consciousness imply our knowledge of their value (LRS 2014: 267; cf. Bramble 2017; Kahane 2016; Lee, n.d.). I canvas the strategy in the remainder of this section, then argue that it fails in §§6-7.

### b. Introspection

Let me begin by pre-empting an objection that I briefly mentioned. Distinguish between the evolution of our *aversion* to pain and our *beliefs* about pain. We evolved a fitness-enhancing mechanism to reliably respond to harmful stimuli, which influenced the formation of the target beliefs. But debunking, the objection goes, doesn't show that believing that pain is bad was adaptive. After all, the belief was not selected in animals that lack propositional attitudes or beliefs altogether, yet many evolved similar mechanisms. Evolution doesn't explain why we believe that pain is bad. At most, it explains our aversion: "The way pleasure and pain feel is already sufficiently motivating." (LRS 2014: 268) Key to the objection is that our beliefs are the product of truth-tracking processes, and so are not vulnerable to the debunking of our aversion

[20] I'm agnostic about Nietzsche's perfectionism but will note that the final value of greatness is compatible with his anti-realism. The latter does put pressure on his commitment to excellences having *intrinsic* value, still they can be *finally* valued (i.e. for their own sake) in a world devoid of mind-independent value.

(Bramble 2017; cf. OWM II 527-8). But as Jaquet (2018: 1154) notes, the cases of masochists and ascetics suggest that "the phenomenologies of pain and pleasure do not always suffice to have us avoid pain and seek pleasure." Moreover, even if our beliefs were a mere by-product of evolved responses, the latter could still have influenced them (1154-5).

This is precisely what the objector denies. For instance, in response to Kahane's claim (2014: 330) that "if *any* normative belief can be given [a debunking] explanation, it is the universal (or near-universal) conviction that pain is bad [for us]", Bramble considers such beliefs to be "*the hardest* to debunk" (2017: 96) and offers an explanation friendly to the hedonist and the objectivist: we evolved an ability to detect the intrinsic quality of pain and pleasure. Our beliefs are "a response to seeing that there is something worth having in pleasure and something worth avoiding in pain." (ibid.) The aversiveness of pain is "built into its very nature" (100).

Similarly, Neil Sinhababu (n.d.; cited by LRS 2014: 267) isolates hedonism from debunking and 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" (n.d: 18) (or pain bad). He assumes the reliability of phenomenal introspection by analogy to visual perception. Our judgments are by-products of a faculty selected for it reliability. Such intuitive judgments then generalize, by virtue of reason: "I should believe that others' pleasures and my pleasures at other times are good" (n.d.: 23) The upshot is two-fold:

(i) *Hedonism*: all and only pleasure (pain) is intrinsically good (bad)

(ii) *Impartiality*: all pleasures (pain) are equally good (bad)

Similar views are at the core of Parfit's theory. Because other moral beliefs are not by-products of similarly reliable faculties, they are susceptible to debunking, but (i) and (ii) are not, the argument goes. That is, our hedonic starting points are not contaminated, evading Street's objection that the reliability of process cannot turn muck into gold. Is phenomenal introspection really a reliable means of access to hedonic facts? Alas, the strategy runs into new problems that the Nietzschean approach illuminates.

## 6. Inner Opacity

The introspective strategy is at odds with what Mattia Riccardi (2015), commenting on Nietzsche, calls "Inner Opacity". It's worth recalling what Nietzsche writes in the Preface to the *Genealogy*:

> We remain of necessity strangers to ourselves, we do not understand ourselves, we *must* mistake ourselves … with respect to ourselves we are not 'knowers'

Part of what Nietzsche means here[21] is that the motives that actually cause our actions are not transparent. Likewise,

> Actions are *never* what they appear to us to be! (D116)

> 'Everyone is farthest from himself' – every person who is expert at scrutinizing the inner life of others knows this to his own chagrin … Your judgment, 'That is right' has a

[21] Surprisingly, the Preface of *Genealogy* is rarely discussed in light of Nietzsche's skepticism about introspection (Gemes 2006; Janaway 2007: 16-19), or superficially so: Katsfanas (2015a) hardly broaches the passage; Riccardi (2015) does not mention it, nor does Leiter (2015), despite citing Wilson (2002). Hanauer (2019) is a recent exception yet says little about the unreliability of introspection. Wilson's (2002) *Strangers to Ourselves* defends a broadly Nietzschean view but does not even once mention Nietzsche.

> prehistory in your drives, inclinations, aversions, experiences, and what you have failed to experience (GS 335)[22]

In GS 354, Nietzsche describes at length the distorting and falsifying tendency of consciousness by means of socially mediated generalizations and language (Katsafanas 2005; Riccardi 2018).[23] In this light, we should expect a mismatch between generalizations about suffering and its experience as a personal source of knowledge. Worse, because consciousness doesn't run deep into our mental lives and often errs, and drives and affects determine our conscious lives, our epistemic access to our subconscious reasons to (not) suffer is both limited and distorted.

To articulate *Inner Opacity*, Riccardi (2015) draws on Schwitzgebel (2008). Nietzsche holds a comparable view (especially D 35, 115; GS 335, 354[24]). So, let us start by considering Schwitzgebel's argument for the unreliability of "naïve" (i.e. untrained[25]) introspection. Because introspection is a multifarious composite of cognitive processes rather than a single unified faculty, our experiences are approached from a variety of facets and by different routes ("there are a hundred ways to listen to your conscience", GS 335). More importantly, whatever information we obtain tends to be inaccurate or ambiguous. Schwitzgebel notes how difficult it is to know whether, say, joy has a single, distinctive, experiential character, and the same generalizes to other affective states.

[22] Also see D 119, 129; GS 333; BGE 32; TI, The Four Great Errors; and Leiter (2015: 81-83).
[23] It is illuminating to read D 114-116 and GS 333, 335, 338 as sequences and in light of GS 354. Shortly after the latter sequence is of course one of the key formulations of eternal recurrence (GS 341).
[24] Riccardi's analysis draws primarily on aphorisms from *Daybreak* and the important GS 354. He does not cite GS 335, which I think plays a pivotal role in articulating Nietzsche's position on the delusions of self-knowledge and the importance of singularity. This aphorism also serves as a transition between the famous passage on Spinoza and the role of drives in the emergence of consciousness (GS 333) and a critique of compassion in light of suffering's personal character (GS 338).
[25] On the mixed prospects of introspective training, see Schwitzgebel (2011: ch. 5).

> Most people are poor introspectors of their own ongoing conscious experience. … We are both ignorant and prone to error. … even in favorable circumstances of careful reflection, with distressing regularity. (2008: 247)

Could pain be an exception (Sinhababu, n.d.: 19)? Schwitzgebel writes about this "favorite example for optimists about introspection":

> 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 notes that Schwitzgebel's purported counterexamples 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." (n.d. 19) Our reflective judgments are indeed prone to confabulation, error and bias, but this hardly helps the hedonist. Our phenomenology would have to be either inherently evaluative or a basis for evaluative judgments, but as we'll see shortly, skepticism is warranted. Even if Schwitzgebel's counterexamples only applied at the level of reflection, phenomenal introspection wouldn't deliver, for we cannot help but reflect on the value of our experiences to bring them to bear on our reasons, and this is where rationalizations occur.

We may, of course, resist Schwitzgebel's interpretation of the evidence. Peter Carruthers (2010) is more sanguine. He assumes there is introspection of perceptual states, yet crucially, he denies there is introspection of *propositional attitudes*—judgments and decisions. Carruthers (2011) develops an "interpretive sensory-access" (ISA) theory of self-knowledge, which I argue also

bolsters *Inner Opacity*. According to ISA, knowledge of one's own thoughts is as interpretive as knowledge of the mental states of others. It draws on the same mindreading faculty, directed at oneself. This faculty cannot but interpret, and only has access to, sensory cues to issue judgments and decisions about others. Self-knowledge draws on similar folk-psychological resources and lacks direct access to our own attitudes, which operate "in the background". Since no other system can give us access to our own attitudes, the mindreading faculty cannot but interpret the available sensory input, including one's physical circumstances, one's and others' behavior, visual imagery, affective feelings, and inner speech. The data, Carruthers argues, reveals that people wrongly attribute thoughts to themselves in response to sensory cues similar to those that would mislead attribution of mental states to others. There is a large body of evidence on confabulation, introspection, and affective forecasting in support of ISA (Gazzaniga 1995; Gilbert et al. 1998; Nisbett and Wilson 1977; Wegner and Wheatley 1999; Wilson 2002). Notably, it turns out people are pretty bad at predicting accurately what will make them happy, by how much, and for how long (Epley 2014: 14-34; Wilson and Gilbert 2003).

Unfortunately, proponents of the introspection strategy tend to presuppose its reliability and have not addressed such skepticism. It is beyond the scope of this paper to assess ISA (cf. Andreotta 2019). Unpersuaded readers can take my argument conditionally: *if* Carruthers' interpretation of the evidence is correct, then we can run an argument from *Inner Opacity* to conclude that we lack epistemically reliable phenomenal access to the badness of suffering. Let me emphasize a few points that, if borne out by the data, would support *Inner Opacity*.

First, the evidence shows how frequently and easily people confabulate about their current or recent thoughts, sincerely self-attributing judgments, goals, or decisions that we know are not

theirs; and not just in reporting the putative causes of their attitudes but also the attitudes themselves. And because they report sincerely and with a sense of immediacy,

> subjects themselves can't tell when they are introspecting and when they are interpreting or confabulating. So, for all we know, it may be that our access to our own judgments and decisions is *always* interpretative, and that we *never* have introspective access to them. (Carruthers 2010: 86).[26]

The idea that we are often deluded about our own motivations (Wilson 2002: ch. 5) is of course central to Nietzsche's conception of agency (see above). The evidence bolsters the Nietzschean approach. We access our preferences, reasons and beliefs through plausible guesses and inferences, observing our own behavior and that of others. But we're no less "strangers to ourselves" than we are to others. We may get lucky, but the same processes we use to infer others' mental states we use on ourselves, and we *know* these processes are not immediate and transparent.

Carruthers' analysis also bears on the present argument because, as ordinary experience reveals, the experience of pain is not detached from salient sensory cues, which individuate the pain as throbbing, stabbing, achy, dull or sharp, as located somewhere, as of a certain duration and intensity, as signaling different sorts of threats, and so on. We also pick up on external cues such as bruises, blood, a dislocated shoulder, and features of the event that caused pain, and how others react to it. Thus, interpretation of various sources of input should influence our self-reported attitudes. More broadly, cueing what others in one's social group respond to suffering

[26] GS 127: "[T]hat a violent stimulus is experienced as pleasure or pain is a matter of the *interpreting* intellect, which, to be sure, generally works without our being conscious of it … and one and the same stimulus *can* be interpreted as pleasure or pain."

and how it is (dis)valued by one's culture, shapes how one interprets the sensory data. For instance, people often say "ouch" *before* experiencing pain, anticipating that, say, bumping one's knee on furniture should hurt – an example of nocebo effect.[27] There is also evidence that our responses to pain and illness are culturally evolved and reliant on social norms. Pain can be turned into pleasure or cancelled by placebos. Culture leads people in some societies to reinterpret pain signals caused by capsicum (found in chili peppers) as excitement or pleasure (Henrich 2016: 11, drawing on work by Paul Rozin and colleagues). We can learn to enjoy the burn and muscle soreness of a good workout. Our cultural expectations about medical treatments can alter our subjective experiences of, and physiological reactions to, pain, even holding underlying chemical stimuli fixed. In sum, culture can override our innate physiological responses (Henrich 2016: 272ff). Thus, our folk theory of what people think about suffering influences our self-attributions, which are interpretative generalizations. And since we have little reason to doubt that most people are similarly positioned, our judgments about suffering result from many justification-undercutting confounds.

Taken together, confabulation and external cueing paint a picture of introspection that is more like narrative construction than it is like archeological discovery. As Wilson and Dunn put it,

> Introspection reveals the *contents* of consciousness … It cannot, however, no matter how deeply people dig, gain direct access to nonconscious mental *processes*. Instead, people must attempt to infer the nature of these processes, by … *constructing* a coherent narrative about themselves (2004: 505; emphasis mine)

[27] Thanks to an anonymous referee for this example.

The reasons why we feel the way we feel and why we do the things we do are not immune. Wilson and Dunn cite "considerable evidence"

> that people have limited access to the reasons for their evaluations and that the process of generating reasons can have negative consequences. … [P]eople do not have complete access to the actual reasons behind their feelings, attitudes, and judgment and thus generate reasons that are consistent with cultural and personal theories and are accessible in memory [Nisbett and Wilson 1977]. But, people do not recognize that the reasons they have just generated are incomplete or inaccurate (ibid.; references omitted)

Central to these findings is that, even when we can access the contents of our thoughts and memories, we are often wrong about the processes underlying the production of our feelings, judgments and behaviors (Nisbett and Wilson 1977; Wilson 2002: 105). This suggests that our beliefs about our phenomenology can often, if not systematically, be the product of justification-undercutting processes. Morever, even if Carruthers admits we can introspect our perceptual states, we still cannot introspect basic normative facts such as the value of suffering, since judging that pain is bad is not a perceptual state. Hence, we should suspend our confidence in the intrinsic reliability of introspection to access the intrinsic badness of suffering. We can sum up the impasse as a dilemma:

(i) *Either the phenomenology of suffering can be explained in terms of (or reduced to) reasons*,[28] but our perceptual states do not explain why we ought to avoid it and our judgments and decisions, where our reasons lie, are not introspectively accessible;

[28] The literature on the nature of pain deals with *motivating* reasons, not reasons in the *normative* sense which interests Parfit. *Evaluativists* explain the unpleasantness of pain by reference to its evaluative content ("this is bad for me") (Helm 2002; Bain 2013; Cutter and Tye 2014); *imperativists* by reference to its imperative content (a command that we avoid it) (Klein 2007; Martínez 2014).

(ii) *Or the phenomenology of suffering cannot be explained in terms of (or reduced to) reasons*, but then phenomenal introspection does not give access to the value of suffering.

A route to our normative reasons presupposes the first horn of the dilemma. But our discussion of introspection casts doubt on the reliability of the judgments we make about our phenomenal experience. Thus, we cannot presuppose the reliability of introspection when arguing for DBS.

Where does this leave us? Not only are there plausible hypotheses about the evolution of our evaluative attitudes, *Inner Opacity* gives us positive reasons to doubt the reliability of the processes by which hedonists and objectivists hoped to withstand debunking. The processes are not just non-truth-tracking; they are inherently falsifying. In the last section, I show that Nietzsche's perspective converges with the empirical evidence.

## 7. Interpretation and generalization

Nietzsche conceives of moral judgments as "symptoms" or "sign-languages" of drives and affects (e.g., BGE 187; D34; 119; 542; TI, Problem, 2; Skirmishes, 37; GM, Preface, 2; cf. Leiter 2013). Moral judgments are caused by, and therefore provide inferential evidence for, underlying drives. For Nietzsche, as noted, claims like DBS are evidence of affective attitudes toward suffering. The diagnosis (synchronic genealogy) casts doubt on their justification by revealing epistemically defective processes.

We are disposed to have certain affective responses as a result of the way our drives are organized. Drives are inherently interpretive and determine our orientation toward the environment, by influencing "perceptual saliences" (Katsafanas 2013, 741), "coloring our view of the world" (743). This framework accounts for our *judgments* because affects have valence:

what we value as expressed by our judgments ultimately reflects what we value unconsciously. The problem is, if affects are noncognitive states, they are not truth-apt, hence not a safe route to DBS.

Other commentators offer a more layered interpretation. 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 problem is now two-fold:

(i) basic affects (e.g. sadness) are not truth-apt; and

(ii) meta-affects (e.g. beliefs about the appropriateness of sadness) are susceptible to confabulation (Knobe and Leiter 2007; Telech and Leiter 2017; D 34, GS 335, BGE 5, KSA 13:14 [116]; cf. §6 above).

Moreover, it is primarily custom that imposes a particular feeling on drives that are, per se, evaluatively neutral (Telech and Leiter 2017; cf. D 38). Accordingly, the underlying affective states of suffering are distinct from the meta-affects, including our judgments about their value. As Telech and Leiter put it, drives are, despite their valence, "morally undetermined". Only our "meta-affective stance (usually culturally shaped, and often involving beliefs) toward the basic affect" constitutes the moral valence of our sentiments (2017: 104). But since we have no reason to think that this stance was shaped by reliable processes skepticism persists.These two interpretations of Nietzsche are compatible; more importantly, both support *Inner Opacity*.

Nietzsche's view of how our conscious moral sentiments are individuated also finds echo in psychologist Lisa Feldman Barrett's (2017) theory that emotions are the variable product of linguistic acculturation and mediated by concepts superimposed on unindividuated affects. The

individuation of affective states under "emotion concepts" (predictive constructed prototypes) does not pick out universal natural kinds with unique fingerprints. The emotions we experience, like sadness or *Schadenfreude*, are shaped by cognition and cultural upbringing. Feldman Barrett also explains why we have concepts to pick out certain states but not others. For the brain, "variation is the norm." (2017: 282) When certain sensations are very intense or very frequent, concepts make sense of our sensory inputs, but what concepts we end up with is contingent. Thus, categorization, relying on statistical regularities and language, shapes our conscious experience of the world.

This resonates with Nietzsche, who cautioned against generalizations. In particular, the distorting role of language is a pervasive theme:

> Language 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)

> man, like every living creature, is constantly thinking but does not know it; the thinking which becomes *conscious* is only the smallest part of it, let's say the shallowest … for only that conscious thinking *takes place in words, that is, in communication symbols* … the development of language and the development of consciousness … go hand in hand. (GS 354)

> We stop valuing ourselves enough when we communicate. Our true experiences are completely taciturn. They could not be communicated even if they wanted to be. … Language, it seems, was invented only for average, mediocre, communicable things. (TI, Skirmishes, 26)

As Katsafanas (2015a: 119) notes, "Nietzsche claims that we lack concepts for most of the mild motives of the sort mentioned … It follows that our ability to bring these motives to consciousness is severely limited." This helps to explain why mixed evaluative experiences such as grief, achievement, masochistic pleasures, or *Schadenfreude,* which all involve unpleasant and pleasant tones, are harder to individuate and categorize. Our concept of suffering does not specify that such experiences are bad. They're ambivalent and the concept is coarse grained. We do experience "anger, hatred, love, pity, desire, knowledge, joy, pain" but these are all "*extreme* states" (D 115), inaccessible in all their contextual and personal nuances. Katsafanas is right that Nietzsche sees our capacity for self-knowledge as limited, but trouble runs even deeper.

We are only introspectively aware of "extreme" or "superlative" states, but our vocabulary is also mediated by social interaction within a certain community (Riccardi 2015). Our self-knowledge is then not just limited but inherently unreliable. Riccardi (2018) also describes the transformation of aggressive drives into bad conscience (profound suffering conceptualized as guilt) through the mediation of the cultural pressures of morality. The "public representation", as Riccardi puts it, that one suffers *because* one is guilty needs to be internalized to shape behavior. Ultimately, the representations we internalize (from priests' sermons to philosophers' platitudes about suffering[29]) become causally efficacious and determine our behavior and conscious lives. Such internalized platitudes include, for instance, that suffering is senseless suffering unless it is deserved (GM III 15) or justified by a god (GM II 7).

Public representations also distort the potential role of suffering in an individual's life. As noted in §3, compassion is not good for most people; it "*strips* of the suffering of what is truly

[29] Cf. GS 326 on the "lies" of the "preachers of morals" about the contributions of pain and misfortune to happiness.

personal" (GS 338). When we seek relief, convalescence, "we want to become estranged from ourself and depersonalized, after pain has for too long and too forcibly made us *personal*" instead of using suffering as the source of knowledge that it is (D 114). The phenomenon is echoed in psychology by "post-traumatic growth" (Joseph 2011). Recall that introspection is fundamentally constructive. Negative events, and subsequent emotions and feelings, can play various roles in the self-narratives we construct, but we can rarely predict what role they will play (Wilson and Gilbert 2003; Wilson 2002, ch. 6). For these reasons, generalizations like DBS are rightly seen with suspicion by Nietzsche. One is not "that which we appear to be in accordance with the states for which alone we have consciousness and word … those cruder outbursts of which alone we are aware make us *misunderstand* ourselves" (D 115). Because suffering's meaning is personal, its concept does not tell us whether or not suffering is bad for us. And by virtue of *Inner Opacity*, we cannot introspect our reasons to approve or disapprove of even particular phenomenal episodes until *second-order* attitudes reinterpret them. Hence, the value of suffering in our lives is not settled by its intrinsic nature, its concept, or what we can phenomenally introspect.

**Conclusion**

In sum, our tendencies to interpret (our phenomenology) and generalize (our experiences and values) explain why most of us believe that suffering is intrinsically bad, even when it is instrumentally necessary. By the same token, these claims explain why this belief is unjustified. We cannot know it by introspection, because phenomenal introspection is not a route to reasons (*Inner Opacity*). But we also cannot know it out of context, because the value of suffering is personal and conferred by meaning-making interpretation as part of a valuable whole (*Narrative value*). These facts support a Nietzschean debunking of a range of views about the value of suffering: realism, hedonism about well-being, and the view that suffering is impersonally and

impartially bad. Nietzsche and cognitive science debunk such beliefs by undercutting their justifications. As we saw, proponents of these views tend to appeal to phenomenal introspection as a last resort against evolutionary debunking. If the present arguments are successful, these views are still in need of justification. This article has also offered a new motivation for Nietzsche's views about the potential value of suffering without committing to specific metaethical position. Instead, Nietzsche's account of susceptibility of consciousness to defective processes dissociates phenomenology from what really matters to us and opens up new potential relations between well-being and suffering.

## References


WORKS BY NIETZSCHE

I use conventional abbreviations, cited by part and section (e.g., GM: II: 3). Posthumous fragments are cited following their classification in the *KSA*, with volume and page numbers.

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

*The Birth of Tragedy and Other Writings*, trans. R. Spiers. Cambridge University Press, 1999)

*Beyond Good and Evil*, trans. J. Norman. Cambridge University Press, 2002

*Daybreak*, trans. R. J. Hollingdale. Cambridge University Press, 1997

*The Gay Science*, trans. J. Nauckhoff and A. del Caro. Cambridge University Press, 2001

*Human, All Too Human*, trans R. J. Hollingdale. Cambridge University Press, 1996

*Kritische Studien-Ausgabe*, ed. 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: Hackett, 1998

*Thus Spoke Zarathustra*, trans. A. Del Caro. Cambridge University Press, 2006

*Untimely Meditations*, trans. R. J. Hollingdale Cambridge University Press, 1983

OTHER WORKS

Andreotta, A. (2019) Confabulation does not undermine introspection for propositional attitudes. *Synthese*. DOI: 10.1007/s11229-019-02373-9

Bain, D. (2013). What makes pains unpleasant? *Philosophical Studies* 166 (1): 69-89

Brady, M. S. (2018) *Suffering and Virtue*. Oxford: Oxford University Press

Bradford, G. (2020) The badness of pain. *Utilitas* 32(2): 236-52

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

Came, D. (2014) Nietzsche on the aesthetics of character and virtue. In D. Came (ed.), *Nietzsche on Art and Life*. Oxford: Oxford University Press, pp. 127-42

Carel, H. and Kidd, I. J. (2020) Suffering as transformative experience. In D. Bain, M. Brady, and J. Corns (eds.), *Philosophy of Suffering: Metaphysics, Value, and Normativity*. London: Routledge, pp. 167-79

Carruthers, P. (2011). *The Opacity of Mind*. New York: Oxford University Press.

—— (2010) Introspection: divided and partly eliminated. *Philosophy & Phenomenological Research* 53(1):76-111

Clark, M. (2012) Suffering and the affirmation of life. *Journal of Nietzsche Studies* 43(1):87-98

—— (2000) Nietzsche's doctrine of the will to power: neither ontological nor biological. *International Studies in Philosophy* 32(3):119-35

Cutter, B. and M. Tye (2011). Tracking representationalism and the painfulness of pain. *Philosophical Issues* 21:90-109

Dancy, J. (2003) Are there organic unities? *Ethics* 113(3):629-50

Delon, N. (2019). Le problème de la souffrance chez Nietzsche et Parfit. *Klêsis* 43:156-186

Epley, N. (2014) *Mindwise*. New York: Knopf

Feldman Barrett, Lisa (2017). *How Emotions are Made: The Secret Life of the Brain.* Mariner Books

Gazzaniga, M. (1995) Consciousness and the cerebral hemispheres. In M. Gazzaniga (ed.), *The Cognitive Neurosciences.* Cambridge: MIT Press, pp. 1391-1400

Gemes, K. (2006) We remain of necessity strangers to ourselves: the key message of Nietzsche's *Genealogy*. In C. D. Acampora (ed.), *Nietzsche's On the Genealogy of Morals: Critical Essays.* Lanham: Rowman & Littlefield, pp. 191-208

Gilbert, D.T., E.C. Pinel, T.D. Wilson, S.J. Blumberg, and T.P. Wheatley (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology* 75(3):617-38

Goldstein, Irwin (1989). Pleasure and pain: unconditional intrinsic values. *Philosophy & Phenomenological Research* 50: 255-76

Hanauer, T. (2019) Strangers to ourselves: self-knowledge in Nietzsche's *Genealogy*. *Journal of Nietzsche Studies* 50(2):250-71

Hassan, P. (forthcoming) Organic unity and the heroic: Nietzsche's aestheticization of suffering. In D. Came (ed.), *Nietzsche on Morality and the Affirmation of Life*. Oxford University Press

Hayward, M. K. (2018) Non-naturalist moral realism and the limits of rational reflection. *Australasian Journal of Philosophy* 96(4):724-37

Helm, B. W. (2002). Felt evaluations: a theory of pleasure and pain. *American Philosophical Quarterly* 39(1):13-30

Henrich, J. (2016). *The Secret of Our Success*. Princeton University Press

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

—— (2014) "Consecration to Culture": Nietzsche on slavery and human dignity. *Journal of the History of Philosophy* 52(1):135-60

Jaquet, F. (2018) Evolution and utilitarianism. *Ethical Theory and Moral Practice* 21:1151–61

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

—— (2007) *Beyond Selflessness: Reading Nietzsche's* Genealogy. Oxford University Press

Joseph, S. (2011) *What Doesn't Kill Us: The New Psychology of Posttraumatic Growth*. New York: Basic Books

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

—— (2001) *The Myth of Morality.* Cambridge: 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):327-41

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

Katsafanas, P. (2015a) Kant and Nietzsche on self-knowledge. In J. Constâncio, M.J. Mayer Branco, and B. Ryan (eds.), *Nietzsche and the Problem of Subjectivity*. Berlin: De Gruyter, pp. 110-30

—— (2015b) Fugitive pleasure and the meaningful Life: Nietzsche on nihilism and higher Values. *Journal of the American Philosophical Association* 1(3):396-416

—— (2013) Nietzsche's philosophical psychology. In K. Gemes and J. Richardson (eds.), *The Oxford Handbook of Nietzsche*. Oxford University Press, pp.727-55

—— (2005) Nietzsche's theory of mind: consciousness and conceptualization. *European Journal of Philosophy* 13:1-31

Kaufman, W. (1974) *Nietzsche: Philosopher, Psychologist, Antichrist*. Princeton University Press

Klein, C. (2014) The Penumbral theory of masochistic pleasure. *Review of Philosophy and Psychology* 5(1):41-55

—— (2007). An imperative theory of pain. *Journal of Philosophy* 104(10):517-32

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

Kraut, R. (2007) *What is Good and Why: The Ethics of Well-Being.* Harvard University Press

Lance, M. and Little, M. (2004) Defeasibility and normative grasp of context. *Erkenntnis* 61:435-55

Lazari-Radek, K. (de) and Singer, P. (2017) Parfit on objectivity and "The profoundest problem of ethics". In P. Singer (2017), pp. 279-96

—— (2014) *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford University Press

Lee, A. Y. (n.d.) How we know pain is bad. Unpublished manuscript. Retrieved at https://www.andrewyuanlee.com/

Leiter, B. (2018) The Truth is Terrible. *Journal of Nietzsche Studies* 49(2):151-73

—— (2015) *Nietzsche on Morality*. Revised edition. New York: Routledge

—— (2013) Moralities are a sign-language of the affects. *Social Philosophy & Policy* 30:237-58

—— (1992) Nietzsche and aestheticism, *Journal of the History of Philosophy* 30(2):275-90

Leknes, S. and Bastian, B. (2014) The benefits of pain. *Review of Philosophy and Psychology* 5:57-70

McPherson, D. (2016) Nietzsche, cosmodicy, and the saintly ideal. *Philosophy* 91(1):39-67

Mollison, J. A. (2018) Nietzsche *contra* stoicism: naturalism and value, suffering and *amor fati*. *Inquiry* 62(1):93-115

Nagel, T. (1986) *The View from Nowhere*. Oxford University Press

Nehamas, A. (1985) *Nietzsche: Life as Literature.* Harvard University Press

Nichols, S. (2014) Process debunking and ethics. *Ethics* 124(4):727-74

—— (2004). *Sentimental Rules*. Oxford: Oxford University Press

Nisbett, R. and Wilson, T. (1977) Telling more than we can know: verbal reports on mental processing. *Psychological Review* 84:231-95

Nussbaum, M. C. (1994) Pity and mercy: Nietzsche's Stoicism. In R. Schacht (ed.), *Nietzsche, Genealogy, Morality*. University of California Press, pp. 138-67

Parfit, D. (1984) *Reasons and Persons*. Oxford University Press

—— (2011) *On What Matters,* 2 volumes. Oxford University Press

—— (2017) *On What Matters,* vol. 3. Oxford University Press

Reginster, B. (2006) *The Affirmation of Life: Nietzsche on Overcoming Nihilism*. Harvard University Press

Riccardi, M. (2018). Nietzsche on the superficiality of consciousness. In M. 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):221-43

Richardson, J. (2015) Nietzsche's Value Monism: Saying Yes to Everything. In M. Dries, and P. J. E. Kail (eds.), *Nietzsche on Mind and Nature*. Oxford University Press, pp. 90-119

—— (2004) *Nietzsche's New Darwinism*. New York: Oxford University Press

Schacht, R. (1983). *Nietzsche*. London: Routledge & Kegan Paul

Schopenhauer, A. (1969) *The World as Will and Representation*, 2 volumes, trans. E. F. J. Payne. New York: Dover

Schwitzgebel, E. (2011) *Perplexities of consciousness*. Cambridge: MIT Press.

—— (2008) The unreliability of naive introspection. *Philosophical Review* 117(2):245-73

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. (n.d.) The epistemic argument for hedonism. Unpublished manuscript. Retrieved at https://philpapers.org/archive/SINTEA-3.pdf

Skarsaune, K. O. (2011). Darwin and moral realism: survival of the iffiest. *Philosophical Studies* 152(2):229-43

Street, S. (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*. Wiley Blackwell

Tersman, F. (2017) Debunking and disagreement. *Noûs* 51(4):754-74

Vavova, K. (2015) Evolutionary debunking of moral realism. *Philosophy Compass* 1(2):104-16

Wegner, D. M. and T. Wheatley (1999) Apparent mental causation. *American Psychologist* 54:480-92

Wilson, T. D. (2002) *Strangers to Ourselves*. Cambridge: Harvard University Press

Wilson, T. D. and D. T. Gilbert (2003). Affective Forecasting. *Advances in Experimental Social Psychology* 35:345-411

Wilson, T. D. and E. W. Dunn (2004). Self-knowledge: its limits, value, and potential for improvement. *Annual Review of Psychology* 55:493-518

Williams, B. (2006) Unbearable suffering. In *The Sense of the Past: Essays in the History of Philosophy*, ed. M. Burnyeat. Princeton University Press, pp. 331-7