Suffering Is Bad

Experiential Understanding and the Impossibility of Intrinsically Valuing Suffering

Louis Gularte

Abstract

Suffering, I argue, is bad. This paper supports that claim by defending a somewhat bolder-sounding one: namely that if anyone—even a sadistic 'amoralist'—fully understands the fact that someone else is suffering, then the only evaluative attitude they can possibly form towards the person's suffering as such is that of being *intrinsically against* it. I first argue that, necessarily, everyone is disposed to be intrinsically against their *own* suffering experiences, *holding fixed their specific overall degree of emotional aversiveness*, because any evaluative attitude other than 'being against'—including mere indifference—would in certain key circumstances make our suffering less emotionally aversive and thus different from the suffering experience (stipulatively) at issue. Second, fully understanding that someone else is having a given experience—Mary's experiencing a vividly blue sky, say, or Job's experiencing heart-rending grief—requires that we represent *experientially* their very instance of that experience-type (it requires, in other words, *token phenomenal concepts*). The result is that what goes for our own suffering goes for others', too: maintaining an accurate experiential representation of the fact that someone else is having a suffering experience with a specific degree of overall emotional aversiveness is only compatible with coming to be intrinsically against their suffering. So suffering is—'objectively'—bad: it's only possible to respond with indifference towards anyone's suffering if we don't fully understand that they are suffering in the first place.

Keywords: suffering, epistemic & practical rationality, understanding, disvalue, phenomenal concepts

Introduction

The claim in my title might not strike you as the discovery of the decade: suffering's badness is likely to seem more obvious than any other normative

fact, and it might even be a *basic* fact, making it a fool's errand to try to argue for it in the first place.

Even so, the objective badness of suffering at least seems to be in tension with another widely-accepted view: namely that it is possible to understand full well that someone else is suffering without being bothered in the slightest. Some people might even fully understand but be positively happy about the other person's suffering, just because that person is suffering. An epistemically idealized version of the reputedly sadistic Roman emperor Caligula, for instance (cf. Gibbard, 1999; Street, 2016), might be exceptionally good at understanding exactly how it feels for his victims to suffer, even in arbitrarily specific ways. But that might be precisely what he enjoys: he might delight in their suffering in proportion to how well he understands, in a vividly experiential way, what he's putting them through.

Such cases challenge the idea that it is irrational or ignorant—*necessarily* irrational or ignorant, for everyone, that is—to fail to care about someone else's suffering. Our imagined Caligula is epistemically perfect, after all (if only by supposition), so maximizing his sadistic joy *might* also be fully practically rational. But if there were nothing practically or epistemically objectionable about Caligula's sadism, then it would be difficult to make sense of the idea that suffering was objectively bad. What would its badness amount to, if it had no objective practical or epistemic 'reason-giving force'—if caring about it and not caring about it could be on equal normative footing?

My goal in this paper is to defend the normative force of suffering's objective badness by arguing directly against the possibility at issue: that is, by arguing that failing to care about someone else's suffering (in contrast to

sadistically valuing it or even just being evaluatively neutral[1]) is only possible if we don't fully understand that they are suffering. In other words:

> If anyone fully understands the fact that someone else is suffering, it will be impossible for them—if they form an evaluative attitude towards it at all—to form any evaluative attitude other than being intrinsically against (intrinsically *disvaluing*[2]) that person's suffering.

If my argument for that claim is sound, then an epistemically idealized, all-understanding, but sadistic Caligula is strictly impossible: practically and epistemically rational or not, if he fully understands the fact that his victims are suffering and forms an evaluative attitude towards that fact, then he will necessarily be disposed to disvalue their suffering.[3] And that means suffering meets a very plausible sufficient condition for being ('objectively') bad (cf. Lewis, 1989).

To be clear, I won't be offering any serious argument in favor of that sufficient condition (although see §3.2), and some people might well reject it. But the possibility of an all-understanding sadist is arguably the core reason for moral skeptics' denial that anything is objectively good or bad. So showing that suffering meets this condition would in effect clear away the main metaethically salient obstacle to accepting that something—namely, suffering—is bad.[4]

---

[1] It may also be possible to neither value, disvalue, nor be evaluatively indifferent to someone's suffering. But the attitudes relevant to suffering's *badness* seem to be precisely those attitudes—namely, the *evaluative* ones. My question is whether, when coming to such an attitude (and thereby 'engaging normatively' in the relevant way), anything rules out intrinsically valuing or being intrinsically evaluatively indifferent towards their suffering.

[2] "Disvaluing"—understood as the negatively-valenced counterpart of "valuing" rather than as equivalent to the privative "*not* valuing" or "*de*valuing"—is more exactly what I mean, but it's less familiar and more technical-sounding, and "being against" is much more common and close enough. I'll be using them interchangeably throughout.

[3] So my paper doesn't, in the end, make any direct claims about failures of epistemic *rationality*: failures of understanding are enough, if something's being 'impossible not to come to disvalue under conditions of full understanding' is sufficient for objective badness.

[4] If you are one of those tempted to reject the sufficient condition at issue, allow me to apologize for my main title, which will no doubt seem too hasty; for you, it should be something like the slightly less catchy "Sadism is impossible under conditions of fully understanding other people's suffering, so if that's sufficient for badness (and it might be), then suffering is bad."

My argument draws on two core claims. The first is that fully understanding the fact that someone else is having an experience of a given type—say, a redness experience, or the taste of cold, fresh-squeezed lemonade—requires that we represent their very instance of that experience-type *experientially* (cf. Jackson, 1986), specifically by means of a *token phenomenal concept*.[5] The second claim is that it's impossible to form any evaluative attitude other than being intrinsically against any experience we represent experientially as *suffering*. Together, those two premises entail that, with respect to the fact that someone else is suffering, it's impossible to come to any evaluative attitude other than being intrinsically against it—as long as we fully understand that they are suffering.

Here is how the paper proceeds. §1 goes into more detail on the key concepts of ‹SUFFERING› and ‹BEING AGAINST›[6] and defends the claim that, for any experience we 'represent experientially as suffering,' it's impossible to form any intrinsic evaluative attitude towards it other than being *intrinsically against* it, by arguing that it's impossible not to disvalue our *own* suffering (if we form an intrinsic evaluative attitude towards it at all). §2 defends the claim that full understanding requires that we represent other people's suffering experientially, by means of token phenomenal concepts, such that we relate to their suffering in a way that is importantly similar to the way we would relate to it if we were experiencing it directly for ourselves. §3 puts those claims together to argue that suffering is — *objectively*, *agent-neutrally*— bad. I close with a brief discussion of possible implications for explaining suffering's badness and its 'reason-giving force.'

[5] See Sundström (2011) for an overview on phenomenal concepts.
[6] I render the thought or concept expressed by "X" as '‹x›.'

1 On being against our own suffering

My goal in §1.1 and §1.2 is to draw out a few intuitive and relatively theory-independent constraints on suffering and disvaluing. The short version of my two proposals is that, first, a given suffering experience is at least partly individuated by the specific degree of overall emotional painfulness or aversiveness (negative affect) of our experience as a whole; and second, if we value or disvalue something, then we are disposed to be, at least in certain privileged contexts, *motivationally* and *affectively* in favor of or against it, respectively.

1.1 On suffering

I'll be arguing, in §1.3, that if we form an evaluative attitude towards our suffering at all, it will be impossible not to be intrinsically against it. A first reason to doubt that claim is that it seems quite possible not to disvalue many experiences we at least wouldn't normally hesitate to *call* 'suffering.' Some people claim to *enjoy* physical pain of certain kinds, for instance, and seemingly intrinsically: just in light of the way such pains feel.[7] And in everyday speech, all or most physical pains could also be described as forms of physical *suffering*.

Still, there is an everyday way of understanding suffering that makes those apparent counterexamples seem like they are missing the point. One way to get into that frame of mind is to note that another natural description of such cases is as instances of pain that don't end up causing us to ('really') suffer—or at least not as much as they would have if we didn't enjoy them.

[7] See Rachels (2003, §4) for skepticism about this possibility. Even more common than enjoying physical pain, probably, is aesthetic appreciation of *emotionally* painful experiences. We sometimes find negative emotions stirred up by music or other forms of art—even deeply, intensely negative emotions—*beautiful* (Smuts, 2007), so it seems possible to value those kinds of suffering, too. I won't actually be arguing against that possibility (see §1.3.2); my claim is just that valuing such experiences would, under certain conditions, make them importantly different experiences (less 'affectively bad-feeling,' in particular) than they otherwise would have been.

Hopefully that way of talking about suffering is intuitive. It may also be worth noting that virtually all philosophical accounts of suffering make the distinction between pain and suffering (Bain, Brady, & Corns, 2020), with suffering usually taken to involve specifically *emotional* pain or 'felt aversion' (cf. Kauppinen, 2020): you only suffer, when experiencing physical pain—or experiencing a loss, annoyance, or anything else—if you 'mind' it, affectively, in the right way.

For that reason, it's probably better not to focus on cases of physical pain, as you read through this paper, but on emotional experiences of, say, grief or sadness. While physical pain may *usually* be emotionally aversive, it also seems to be the sort of thing we can in principle enjoy—or at least experience in an affectively neutral way as a result of, say, meditation practice (Grant, 2014) or brain injury (Grahek, 2007). And that could conceivably make it less emotionally aversive[8] without making it any less physically *painful*.

As I hinted above, then, the minimal feature of suffering I'll be drawing on—a feature compatible with virtually all accounts of experiential suffering—is that each suffering experience comes with a specific degree of overall emotional aversiveness ('overall affective negativeness'). The most paradigmatic, thoroughgoing instances of suffering, for instance, seem to involve a kind of *unmitigated* overall emotional aversiveness.[9]

An important consequence, for the sections to follow, is that when I talk about suffering experiences, what I have in mind are *whole* experiences—almost always with some emotionally neutral or positive components in

[8] Pain asymbolia (Grahek, 2007), in particular, can reportedly make the experience of pain entirely free of emotional unpleasantness.

[9] There is precedent for understanding suffering itself—rather than just the 'most suffering-y' suffering—in something like that way: some have thought of suffering as negative emotion towards *our situation as a whole* (Kauppinen, 2020; see also (though with many caveats) "Sallasutta: An Arrow", 2018) or with the frustration not of mere desires or wishes but of our resolute *will* (Schopenhauer, 2010, e.g., §65).

addition to the negative ones—rather than just the experience's emotionally negative *parts*.[10] The reason is just that that's how actual suffering experiences are. We don't experience the negative and positive components in isolation; we experience them together, making every given suffering experience come with a specific degree of *overall* emotional aversiveness.

1.2 On 'being against'

Turning to the attitude of 'being against' or *disvaluing*, the main question is whether the kind of attitudes at issue—valuing, disvaluing, and evaluative neutrality—are cognitive, non-cognitive, something in between, or some combination of the two.

All that matters for the argument to come, though, is that whatever cognitive features they may have, attitudes of valuing, disvaluing, and evaluative neutrality seemingly necessarily involve, at least under certain specific conditions, some corresponding affect and motivation. If someone claims to be against something—racism, say—but never feels or would feel even the slightest negative affect or motivation of any kind against it,[11] then it's natural to think they don't really disvalue racism.

It *might* still be possible (though it also might not be: cf. Rosati, 2016) for someone who is 'left cold' by racism in that way to make a related sincere *judgment* or form a related normative belief, such as ‹it's good to combat racism›. But we would be justified in characterizing such a person—someone who truly had no accompanying motivating or affective attitude in favor of combating racism—as someone who didn't care about combating racism in the way required for it to count among their values, properly understood.

---

[10] Thanks to an anonymous reviewer for pressing me to clarify this point.

[11] I'm understanding affect and motivation quite broadly, to include what we could call 'meta-affect' and 'meta-motivation': motivation, say, to develop (stronger) first-order motivation.

In short, it seems safe to endorse a very weak version of affective and motivational *internalism* about valuing and disvaluing: valuing something, for instance, entails that at least under certain specific conditions (see §1.3.1) and to some subtle degree, we would feel affectively positive towards, and motivated in favor of, what we valued.[12]

Before I move on, note that on its own, §1.1's minimal conceptual constraint on suffering—that each suffering experience comes with a specific degree of overall emotional aversiveness—has no direct entailments for the attitude of disvaluing as I'm understanding it here; my argument below is not a defense of the tautology that we're necessarily disposed to disvalue the kind of suffering that we're necessarily disposed to disvalue. Even if there turns out to be *some* tight relationship between them, overall emotional aversiveness is neither, say, *constituted* by nor conceptually inseparable from disvaluing,[13] which may well include (for all I've said here) much more complicated cognitive attitudes. Overall emotional aversiveness is an experiential quality with which we are all familiar and which we can understand independently of disvaluing.

1.3 Evaluative attitudes towards our own suffering

Together, those two minimal constraints on suffering and valuing provide an argument for my first main claim: necessarily, everyone is disposed to be intrinsically against their own suffering—or any experience they 'represent experientially as suffering.'

[12] You might wonder why I'm including both affect and motivation rather than just one of them. The reason is that feeling motivated but never to any degree affectively 'bothered' one way or the other, if that's even possible (Scarantino, 2014)—and *vice versa*—again seem to entail that we don't *care* in the way we care when we value or disvalue something.

[13] I'm not hereby ruling out evaluativist accounts of aversiveness or painfulness (e.g. Bain, 2017), on which they are constituted by a representation that something is bad; disvaluing seems related to but distinct from representing something as bad as such.

Here is the core idea, without further ado: if someone comes to value or be evaluatively neutral towards their suffering, then their overall experience will—at least in certain key moments—thereby become less overall emotionally aversive. In slogan form, *if you're into it, then you're not suffering as much*. The result is that it's impossible to come to value (positively or neutrally) the specific 'pre-evaluative-attitude' suffering at issue while maintaining its specific degree of overall emotional aversiveness.

1.3.1 The 'test conditions' for intrinsically valuing our suffering

It's worth addressing one potential worry right away. Remember that on my proposed way of understanding valuing (§1.2), it only necessarily involves feeling affect and motivation *under certain specific conditions*. So it might sound possible, despite the argument I just introduced, to come to value a suffering experience intrinsically without thereby reducing its overall emotional aversiveness. In particular, as long as the 'certain specific conditions' in which we have the relevant affective attitude (associated with the evaluative attitude) only come about when we're not actively experiencing our suffering, the argument sketched above won't secure the impossibility of coming to value, intrinsically, our suffering: there would never be any conflict between our suffering experience's degree of overall emotional aversiveness and the positive affective contribution of our valuing it.

It's true that §1.2's proposal—that valuing requires that we would feel the corresponding affective attitude, at least under certain key conditions—left that possibility open. The thought was just that having no positive or negative affective disposition whatsoever about something would entail that we neither valued nor disvalued it. Valuing and disvaluing (again, possibly in contrast to attitudes like simply *believing something is valuable*) get us 'affectively involved.' So I now want to make two specific

points in connection with the conditions in which our evaluative attitudes would 'contribute affectively,' to help show why coming to value our suffering experiences as such—holding fixed their specific degrees of overall emotional aversiveness—is impossible.

First, note that if I'm right and anyone who values something positively, negatively, or neutrally is necessarily disposed to have the corresponding affective attitude under *some* conditions, then the *test* of whether they intrinsically value something—friendship, say—concerns the moments in which they are or would be experiencing the relevant affective attitude characteristic of valuing (an affectively *positive* attitude, at minimum) towards friendship. They only intrinsically value friendship, in other words, if there are or would be such moments of being affectively involved in the way associated with valuing it.

Second, note that we are specifically interested in what follows from valuing something *intrinsically*. As I use that expression in this paper, to value something intrinsically is to value it *just in light of that 'something' itself*: just in light of its essential or constitutive features, or in light of understanding it for what it is.[14] But suffering, as I'm understanding it (see §1.1), is essentially experiential: one of its essential features is its *feeling the way it feels*,

[14]The expression is probably infelicitous for *dis*valuing, but another common and related way of glossing *valuing intrinsically* is valuing something "for its own sake" (e.g. Arpaly & Schroeder, 2013, §I.3). Note that "intrinsic (dis)valuing" can be used less narrowly than the way I'll be using it. As I'm understanding them here, for instance, intrinsic valuing and disvaluing involve more than just *non-instrumental* (and non-'realizer' (Arpaly & Schroeder, 2013, §I.3)) valuing and disvaluing.

Note also that, in keeping with my neutrality on the nature of (dis)valuing as such (§1.2), the way I'm using "intrinsic valuing" is not meant to rule out Humean views of emotions or valuing (or both) according to which they are 'original existences' with no representational content of their own (e.g. Whiting, 2020). On such views, someone *S*'s valuing some *X* 'just in light of its essential or constitutive features' should just be understood as *S*'s having the relevant evaluative or affective mental component *M*, where *M* stands in ('Humean') relation *H* to representation *R*, such that (i) *R* is *S*'s representation of '*X* as such,' or *X*'s essential properties, or the properties of *X* mentally but 'constantly' associated with *X* 'taken on its own'; and (ii) Humean relation *H* is specified by the relevant preferred Humean account of the affect–representation relationship (e.g., *M* "follows on immediately from [and] is directly caused by" *R* (Whiting, 2011, 288)). Thanks to an anonymous reviewer for inviting me to address this question.

experientially.[15] Coming to an evaluative attitude towards it just in light of its essential features thus requires that we be in conditions of experientially appreciating 'what it's like' to have the relevant suffering experience.

The result is that another part of the 'test' of whether someone values their own suffering *intrinsically* is whether they value their suffering when confronted with it in its full experiential glory—that is, while experiencing it and focusing just on its precise constitutive experiential qualities (cf. §2)

Putting those observations together gives the following result for the conditions of the 'test' of whether someone intrinsically positively, neutrally, or negatively values their suffering. In particular, if a person values their suffering, then there must be circumstances in which they would be both (i) experiencing that suffering, with its same specific degree of overall emotional aversiveness, while also (ii) coming to experience the positive, negative, or neutral affect (towards that suffering) that characteristically accompanies valuing, disvaluing, or being evaluatively neutral, respectively. That is, someone only intrinsically values their own suffering if they could come to feel affectively positive towards their suffering experience as a whole *while actively experiencing it*.

It follows that coming to value (positively or neutrally) our suffering experience as a whole, with its specific degree of overall emotional aversiveness—coming to value it *intrinsically*, that is—is impossible: it would only be possible if it were possible under those two conditions: (i) actively experiencing the suffering and (ii) experiencing the new valuing-specific affect; and it isn't. The reason, as I introduced above, is that adding positive

[15]Some accounts of suffering reject this 'essentially experiential' commitment (e.g. Corns, 2021; Carruthers, 2004; for a defense of experiential accounts, see Coleman, 2019). If they're right, my paper should be taken to be concerned with one specifically experiential subtype of suffering, identified by the constraint proposed in §1.1.

or neutral affect to our suffering experience reduces its degree of overall emotional aversiveness, thereby constitutively chasing away the *specific* degree of overall emotional aversiveness that was in question in the first place.

1.3.2 Other ways of valuing our suffering

It's worth emphasizing that this argument says nothing about the possibility of coming to value our suffering experience *non*-intrinsically—as, say, a necessary step towards some larger goal, or even just as a sign of progress towards it. First, the 'test' of non-intrinsic valuing doesn't require that there be moments of confronting suffering 'in its full experiential glory' at all. What makes us count as (say) instrumentally valuing something is just our attitude towards the relevant *instrumental* features. And second, even if there were moments in which instrumentally valuing our suffering made our overall experience less emotionally aversive,[16] we might still count as valuing our suffering—just as long as by "our suffering" we don't mean our specific suffering experience as a whole, *with its specific previous degree of overall emotional aversiveness*. In fact, the same goes for coming to value our suffering intrinsically: if we use "our suffering" that way—to pick out, say, a specific affectively negative experience-*component* (see §1.1)—then it's perfectly possible to come to value 'our suffering' intrinsically.

For parallel reasons, on yet another way of understanding the expression "our suffering," my argument doesn't even rule out the possibility of coming to value our suffering intrinsically *while* holding fixed the degree of emotional aversiveness we are experiencing over time. In particular, that will be possible if by "our suffering" we mean not some *specific* 'pre-evaluative

[16] We may even have good instrumental reasons not just to suffer but to *value* our suffering, in some contexts (cf. McRae, 2018; Stump, 2010), and one of those reasons could be that valuing it would make our suffering less emotionally aversive. For an eager denunciation of this particular emotion-regulation strategy, see Nietzsche (1887).

attitude' suffering experience, but *any* experience with the same degree of overall emotional aversiveness.

Here's an example. Suppose that, at t_1, we are experiencing (i) the emotional aversiveness of a stinging insult about our choice of hairstyles but also (ii) the emotional pleasingness of a tasty lollipop in our mouth, with a resulting overall affective aversiveness-to-pleasingness ratio of (say) 3:1. Now suppose that, at t_2, we have managed to maintain exactly the same 3:1 aversiveness-to-pleasingness ratio but have *replaced* the emotional pleasingness contributed by the lollipop with the emotional pleasingness contributed by *valuing this* (new) *suffering experience*—that is, the experience constituted by the aversiveness of the insult and the pleasingness associated with valuing the new but still 3:1 aversive-to-pleasing experience as a whole. We would then have come to value a suffering experience while maintaining the same degree of overall emotional aversiveness.

Note, though, that the main question relevant to generalizing from the case of our own suffering to the suffering of others is what *new* evaluative attitudes we can form towards a given pre-existing, specific suffering experience as a whole, with its specific degree of overall emotional aversiveness. The 'valuing-replaces-lollipop' case does involve the formation of a new evaluative attitude,[17] but that new attitude is also towards a new and importantly different suffering experience. It is partly constituted by our now *liking* our experience *qua* suffering experience, after all, and has no lollipop-based pleasingness. So we can ask my question with respect to *either* suffering experience in the 'valuing-replaces-lollipop' case: What new

[17] A close cousin of the 'valuing by replacement' case is the possibility, perhaps thanks to a lifetime of Buddhist Lojong mind-training practice (e.g. Atisha, 2017), of never experiencing any suffering without *at the same time* valuing—and experiencing a positive appreciative feeling towards—that suffering. My argument by no means rules such cases out; it's just that (as in the main text), *with respect to generalizing to the interpersonal case*, the relevant question would be what new evaluative attitudes the possible Buddhist monk at issue might form towards her original suffering-plus-valuing experience without decreasing *that* experience's degree of overall emotional aversiveness.

evaluative attitudes can we form towards the insult-plus-lollipop suffering experience? What about towards the insult-plus-valuing one? In both cases, the argument in §1.3.1. entails that we can only come to disvalue them. The mere existence of the insult-plus-valuing experience, then, doesn't constitute an answer to the question of what attitudes we can form towards the insult-plus-*lollipop* experience, despite their sharing the same degree of overall emotional aversiveness, and despite, in the insult-plus-valuing case, our valuing our suffering experience in light of its degree of overall emotional aversiveness.

More generally, the point of the argument in this section is just that given any particular suffering experience—a specific experience of suffering, with a specific degree of overall emotional aversiveness—coming to value it positively or neutrally would make the overall suffering experience we were experiencing less emotionally aversive, in the relevant 'test' conditions, and thus an importantly different experience.

It would be importantly different, that is, with respect to the standard of *having exactly the same degree of overall emotional aversiveness*—even if it were in all other respects the same. That standard may not, in general, be particularly practically relevant when it comes to our own suffering: why not make our suffering less emotionally aversive? But again, and as I will be arguing in the sections to come, it turns out to be a central epistemic requirement for understanding someone *else*'s specific suffering experience as a whole, which (unlike our own suffering) doesn't change, experientially—let alone for the better—just in virtue of our coming to value it.

1.3.3 'Intrinsic meta-affect' and the possibility of disvaluing our suffering

A last question concerns the affective contribution of *dis*valuing. All I have said so far is that, *in the relevant* ('test') *conditions*, valuing and evaluative

neutrality would lessen our experience's overall emotional aversiveness and thus make it an importantly different suffering experience from the one we were supposedly forming the attitude towards. But you might think, for parallel reasons, that *dis*valuing our suffering would *increase* its emotional aversiveness. If that were so, it would be impossible to come to any evaluative attitude whatsoever with respect to our own suffering's specific pre-evaluative-attitude degree of overall emotional aversiveness—disvaluing included.

My answer is that, as others have observed (e.g. Klein, 2015, 55; Korsgaard, 1996, 154), it seems to be an intrinsic feature of suffering that we are averse not only to the subjective source of our suffering but to our suffering itself—at least when focusing on it, as we're here supposing we are. When we're suffering, then, we already have an affectively negative attitude towards our suffering: 'intrinsic meta-aversion' is an existing part of our suffering experience's degree of overall emotional aversiveness.[18]

The result is that we can come to disvalue our suffering intrinsically without increasing its degree of overall emotional aversiveness. Suffering's existing intrinsic meta-aversion can *be* (or be 'recruited' or 'co-opted' as) the affective component of our disvaluing. Intrinsic meta-aversion is directed at our suffering, after all, and in all likelihood our intrinsically generated meta-aversion to suffering is in fact part of the unpleasantness of many unpleasant everyday experiences.

[18] A parallel phenomenon no doubt holds for affectively positive experiences: at least when we're focusing on how they feel (and not on, say, their fleetingness), it's arguably part of the nature of happiness experiences to be partly constituted by 'intrinsic meta-liking.' So it should be possible, for parallel reasons, to come to value our happiness intrinsically without increasing its overall degree of emotional pleasingness: intrinsic meta-liking can act as the affective component of our valuing. An important consequence is that the argument of this section also doesn't rule out valuing a given affective state—even a suffering state—in light of its having whatever non-zero degree of overall *positive* affect it has. My claim is just that *qua suffering*-state (with its specific degree of overall emotional aversiveness), the only evaluative attitude towards it that wouldn't reduce our experience's overall emotional aversiveness, in the relevant conditions, is disvaluing.

In sum, the only evaluative attitude compatible with sustaining a given suffering experience as a whole—with its specific degree of overall emotional aversiveness—is disvaluing it. So, necessarily, everyone is disposed to be intrinsically against their own suffering.[19]

2 Full understanding and token phenomenal concepts

It may already have sounded intuitive to you that necessarily everyone is disposed to disvalue their own suffering (if they form an evaluative attitude towards it at all)—and on its own, that fact doesn't entail anything about what attitudes we can form towards *other* people's suffering. So you might take what I've said so far to establish that suffering is bad 'for the sufferer'; but Caligula, for instance, might happily agree, and just take it as further fuel for his sadistic delight: he *likes* that suffering is bad for his victims.

So this section moves on to defend the second main premise of the paper: namely, that fully understanding the fact that someone else is suffering requires that we represent their suffering experientially. If so, then (by the argument in §1.3), necessarily, anyone—Caligula included—who fully understood the fact that someone else was suffering could only come to be intrinsically against it.

A few important clarifications are in order before diving in. A first point to acknowledge is that some things are no doubt just impossible to understand, fully or otherwise, and for all I've said so far, that could turn out to be the case for understanding *that someone else is suffering*. If so, that might make

[19]There are plenty of apparent counterexamples, to be sure: religious ascetics, for instance, or perhaps wholehearted retributivists (cf. note 50), or just people racked by guilt and committed to the idea that they deserve to suffer, might at least take themselves to value their suffering intrinsically—even their *maximally overall emotionally aversive* suffering. As I already noted in the main text, my proposal is compatible with valuing suffering: intrinsically valuing our experience's being *to some unspecified degree or other* emotionally aversive doesn't mean our experience as a whole loses all overall emotional aversiveness; it just reduces it. But my argument does force me to deny that we can value—intrinsically and thus while fully in the grips of the experience—our own *maximally* emotionally aversive suffering experiences.

it hard to interpret my claim—that anyone who fully understood that someone was suffering could only come to be intrinsically against that fact—as establishing a practical-rationality-relevant epistemic deficiency in everyone who intrinsically valued or was evaluatively indifferent about someone else's suffering. In particular, how could it establish a necessary epistemic deficiency if someone like Caligula, say, could go on intrinsically valuing someone's suffering even when understanding *as fully as humanly possible* that the person was suffering?[20]

Now, the most straightforward barriers to full understanding are the finite time, storage, and 'processing power' available to our minds: some things might just be too complex, or require knowledge of too much information, to afford full understanding by a given individual, or even by humans in general. It's also arguably impossible for *any* mind—finite or not—to understand fully something like *what it's like to be devoting 100% of one's attention to (say) the taste of their coffee* while at the same exact time also fully understanding what it's like to be devoting just *half* of one's attention to the taste of their coffee.[21]

In order to see whether such worries apply to our case, the main thing to note is that the question my argument is concerned with is (i) which *intrinsic* evaluative attitudes we can form (ii) towards a fact constituted by a very minimal set of features: namely, the fact that *someone other than ourselves* is *suffering*. Next, recall that on my understanding of intrinsic valuing,

---

[20]Note, in contrast, that showing that it *is* humanly possible to have full understanding of the bare fact that someone else is suffering (in conjunction with establishing that full understanding is incompatible with forming or maintaining an attitude of evaluative indifference) would indeed seem to be enough, intuitively, to establish the relevant guaranteed epistemic deficiency in all alleged amoralists—that is, even if a given *specific* individual lacks (for, say, neurological reasons) the necessary cognitive or affective faculties.

[21]Thanks to an anonymous reviewer for drawing my attention to these worries. Note that 'finiteness'-related barriers to full understanding would entail the impossibility of fully understanding *anything*, for humans, if fully understanding something required perfect knowledge of its infinite so-called 'Cambridge properties' (see (e.g.) Francescotti, 1999) — properties like *being such that 1 is one more than 0, being such that* 4^7 *is three more than the 1,900*th *prime number & the average elevation of the Netherlands is approximately 30m*, . . . and so on. For reasons I'm about to lay out in the main text, such properties are nonetheless very unlikely to be relevant to the kind of understanding at issue in my argument.

it involves valuing something 'just in light of its essential or constitutive properties' (§1.3.1). The result is that counting as having a given attitude *intrinsically* towards the fact that someone else is suffering requires fully understanding just (i) that there is someone who isn't *us* and (ii) that they (whoever they are) are experiencing an instance of the experience-type *suffering*. Finally, if my observations in the sections to come are right, then it should in fact be possible to understand—even fully—those much more limited features of the suffering-situation at issue.

One last point to note is that even though I'll be defending in passing the idea that fully understanding those minimal features is possible, my argument actually doesn't hinge on whether that's true. The reason is that the core relevant premise of my argument really only concerns a specific *necessary* condition for fully understanding that someone is suffering—namely, again, that we represent the other person's suffering *experientially*, with an experiential mode of presentation or 'token phenomenal concept.'

Strictly speaking, then, as long as my argument for the possibility of satisfying that necessary condition works, my central claim could be restated in terms of the "fullest *(humanly) possible* understanding": necessarily, anyone who understood as fully as (humanly) possible the bare fact that someone else was suffering, to a given specific degree of emotional aversiveness, could only come to disvalue, intrinsically, that person's suffering.

2.1 On 'understanding'

Quite a few historical and contemporary metaethical projects have tried to establish something like that claim: namely, that having a certain epistemic good would make it impossible not to care about other people or their well-being (e.g. Kant, 2002; Nagel, 1970; possibly even Hume: see Railton, 1999). But the relevant 'epistemic goods' vary. The projects closest to mine, for

instance, are Colin Marshall's (2018) *Compassionate Moral Realism* and Emad Atiq's (2021) "Acquaintance, knowledge, and value" (cf. Atiq & Duncan, in press; Sinhababu, 2022). Marshall's target epistemic good is *being in touch*, and Atiq's *(acquaintance) knowledge*.[22]

There are no doubt important differences between the target goods, but my 'full understanding' is meant to be maximally inclusive: my concern is with epistemic improvements of any kind, such that *fully* understanding something entails that no epistemic situation with respect to it would be (epistemically) better. My actual target is thus the broad idea corresponding to—to use a colloquial expression—(as fully as possible) *knowing what we're talking about*.

The reason for my maximally inclusive notion of epistemic value is that, first, when it comes to representing experiences or experiential qualities—the taste of fresh-squeezed lemonade, say—there is already quite a broad consensus (see Nida-Rümelin & O'Conaill, 2021) that someone who has never had a given kind of experience is importantly epistemically impoverished with respect to that experience-type.[23] And second, the core reasons motivating that consensus are ultimately all my argument relies on.

---

[22]The more important difference between my argument and Marshall's and Atiq's has to do with the possibility of the 'epistemically perfect but sadistic Caligula,' who we are to suppose is perfectly in touch with, has perfect acquaintance knowledge of, fully understands . . . (and so on) his victim's suffering experience—including how bad it feels—but who is also very much in favor of that situation as a whole. Unlike Atiq's and Marshall's arguments, my response draws centrally on observations about the *degree* of emotional aversiveness of a given sufferer's suffering. In particular, my argument's focus is on fully understanding the sufferer's suffering experience in light of its degree of overall emotional aversiveness, and the implications for the (specifically *evaluative*) attitudes Caligula can form towards the sufferer's suffering experience as a whole.

While my approach to the problem is thus (to my knowledge) novel, I nonetheless see it as fleshing out a rough strategy Marshall and Atiq both seem to find attractive. Atiq, for instance, doesn't directly address Caligula-type cases, but he does claim that "being acquainted with the sensible content of others' suffering [entails that] one feels aversion towards the pains of others *just as they do*" (Atiq, 2021, §5, my emphasis); and Marshall argues that "the most obviously morally problematic conflicted sadists are not *proportionately* in touch with the pain they cause[:] though the pain they cause [. . .] is greater than the pleasure they experience, they are more pleased by their own pleasure than pained by others' pain" (Marshall, 2018, 140). So my argument can be seen as taking up where Atiq and Marshall left off, developing a somewhat more precise version of a possible through-line on display in such passages.

[23]Note that being 'epistemically impoverished' needn't involve any lack of *propositional knowledge*. Lewis (1990), for instance, agrees that there is an epistemic improvement attainable through direct experience (namely, the improvement constituted by now "knowing what [the] experience is like" (Lewis, 1990, 517)), but he characterizes the improvement as a form of 'knowing-how' rather than 'knowing-that' (cf. Nida-Rümelin & O'Conaill, 2021, §4.3 and my comments on *understanding* in the main text).

It's worth noting, despite my very inclusive notion of epistemic value, that most dedicated theories of *understanding* as such would also support categorizing the specific epistemic good at issue in this section as (at least) a form of understanding. First, understanding a given fact is usually taken to be less compatible with epistemic deference than merely *knowing* the same fact (e.g. Hills, 2016).[24] And second, while some have argued that understanding might ultimately reduce to propositional knowledge (e.g. Sliwa, 2015; cf. Lawler, 2018), even they think that, relative to just knowing some fact, understanding it generally requires richer information and more fine-tuned cognitive abilities (Grimm, 2021; Pritchard, 2009, §2.2) or 'know-how' (cf. Hills, 2016).

What follows for understanding experiential qualities like suffering, given those established commitments about understanding? A natural proposal is that fully understanding an experience-type requires the specific 'know how,' gained from experiencing it for ourselves, of being able to *recognize* the target experience-type—say, the taste of cold, fresh-squeezed lemonade, again—when experiencing it (cf. Lewis, 1990; Sliwa, 2017).

We can go further, though: in particular, *being actively confronted* with the experience also seems to be doing some of the epistemic work in these understanding-related moments of recognition.

Consider Jackson's (1986) famous example of Mary the color scientist who has lived in a gray-scale environment her whole life and finally sees, say, a stunningly blue sky. Now imagine that Mary returns to her gray-scale environment and completely forgets how the sky looked—despite, we can suppose, now having the ability to recognize blueness if she sees it again. It seems clear that Mary is in a worse epistemic position, blueness-wise, than

[24]The parallel asymmetry might not hold for knowledge of *things* (Duncan, 2020; cf. Atiq, 2021; Atiq & Duncan, in press) — or for knowledge as implicitly understood in the everyday notion of 'knowing what we're talking about.'

when she was outside. What's more, if she is suddenly struck by a vivid visual memory of the blue sky, it seems clear that she is back in a better epistemic position.

Arguably for the same reason, Mary also seems to be in a worse epistemic position with respect to blueness whenever she isn't actively recalling (or seeing) something blue, even if she could summon up a vivid memory on command. We might not hesitate to *say*—though on my view we would be speaking somewhat loosely—that Mary, understood as a temporally extended thinker and imaginer, 'knows full well what she's talking about,' now, when talking about blueness experiences. And yet, as Emad Atiq (Atiq, 2021, §3) has argued at greater length, Mary remains importantly like her former, experiential-blueness-benighted self, in these specific moments: she could, for instance, coherently *wonder* whether she has completely forgotten what blueness looks like or whether she can still simulate a blue mental image on command—even if she *in fact* can. Relatedly, the moments actually constituting her remembering what a blue sky looks like are, arguably, the moments in which she is actively experiencing a visual memory of a blue sky. So there are good reasons to think that it's only in those moments that Mary is in the best epistemic position with respect to experiential blueness.[25]

[25]Note that neither Atiq (as I read him) nor I should be taken as attempting to *explain* (rather than just argue for) the irreplaceable epistemic value of 'actively experiential' representation. My own suspicion is that a successful explanation would draw on at least the following points:

(i) experiential representation directly and experientially *exemplifies* (e.g. Elgin, 1983; Goodman, 1976) the relevant experiential qualities, whereas all other ways of representing them do so *indirectly* (say, 'by description,' using lexical concepts);

(ii) full understanding requires directly mentally instantiating (at least) the relevant *structural relations* of the fact or phenomenon to be understood (cf. Strevens, 2008; Zagzebski, 2019);

(iii) experiences of experiential qualities are at least partly constituted by *essentially perspectival* structural relations (cf. Nagel, 1974).

The simplest motivation for the third point is the observation that instances of experiential qualities necessarily stand in the perspectival relation of *being experienced by an experiencer*. An attractive second reason to accept it would be a version of structuralism about the very nature of each experiential quality as such (cf. Clark, 2000; Lyre, 2022), which may be especially appealing in the case of suffering (cf. Clark, 2005) even if you think it flounders for (e.g.) color-qualia. In any case, [(i)-(iii)], if true, would suggest that, in the case of experiential qualities, only experiential representation directly mentally instantiates the relevant structural relations, so only experiential representation allows for full understanding. Thanks to Dan Burnston for the suggestion about exemplification.

In any case, I will hereafter be assuming that when it comes to something essentially experiential like suffering, we are always in a better epistemic position to the degree that we are actively experiencing (or vividly 'simulating'[26] or recalling) the relevant experiential quality. For anyone unswayed by the above observations or related arguments from Atiq and others, my overall claim in this paper should thus be understood conditionally: *if* fully understanding something essentially experiential requires an 'actively experiential' mode of presentation, then suffering is bad.

To be sure, that is more understanding than we need, for most purposes. The point is that any evaluative attitudes—towards, say, the color blue or the taste of lemonade—that we would form if we *did* have this more robust kind of understanding would be more authentically about blueness or lemonade-taste *as they really are*.

In any case, as promised, this proposed requirement for full understanding is equivalent to one requiring that we represent things like blueness or the taste of lemonade using phenomenal concepts: here *type* phenomenal concepts, and specifically ones constituted by an occurrent instance[27]—understood broadly to include vivid memories or mental imagery—of an experience of the relevant type.

There is an active debate over the nature of phenomenal concepts. Some (e.g. Levin, 2006) say they are are best understood as being like *demonstratives*: deploying a phenomenal concept of the taste of lemonade is like thinking ‹IT TASTES LIKE ð THISÞ›, where ‹ð THISÞ› directly refers to the remembered, simulated, or actual taste-experience or the experience-type it

[26]Though I won't be discussing empathy (see Bailey, 2022 for a project parallel to mine but with empathy as its focus), my proposal about what is required for full understanding has some things in common with the 'simulation' theory of empathy (for an overview, see Barlassina & Gordon, 2017). See Mayerfeld (1999, §4.4) for related observations on the specific case of understanding suffering.

[27]Type phenomenal concepts may also be possible without an occurrent instance of the relevant experience-type (cf. Balog, 2009; Chalmers, 2003, §2; for more on related thinly recognitional accounts of understanding, see Levin, 2006; Loar, 1997; Sosa, 2003), but given my argument in the main text, I'll hereafter use "phenomenal concepts" just for ones constituted by an occurrent instance.

instantiates.[28] Katalin Balog's (2012) 'quotational' account, in contrast, proposes a "concept-forming mechanism that operates on an experience and turns it into a phenomenal concept that refers to either the token experience, or to a type of phenomenal experience that the token exemplifies" (Balog, 2012, 33).

For my purposes, all that matters is that phenomenal concepts can be taken to constitute an essentially experiential representational format or mode of presentation.[29] Representing the taste of cold, fresh-squeezed lemonade using a phenomenal concept makes us relate to that taste-type in an experiential way, partly constituted by our currently (re)experiencing the lemonade-specific experiential character.

2.2 Token phenomenal concepts of other people's experiences

In short, fully understanding an experience-type seems to require that we represent it experientially. Now, that might sound like it already establishes what I set out to establish: if understanding the type *suffering* requires that we represent it experientially, then the argument in §1.3 might entail that (if they come to an evaluative attitude about it at all) it's impossible for anyone not to disvalue not just their own suffering but the whole experience-type of which it is an instance.

Even if that were true, though, it would be unclear what followed from it. What is it to disvalue an experience-type? If it entails disvaluing all instances of the type ('in' anyone, ever), then I don't need any further argument for the conclusion that necessarily, anyone who fully understood the fact that someone else was suffering would only be able to come to be intrinsically

[28] I'll be using the 'demonstrative'-style notation throughout (my '⟨ϕ⟩' brackets together make a "Φ" for *phenomenal*)—but just for ease of expression: I will be staying neutral on the debate in the main text.

[29] 'Demonstrative' theorists tend to be leery about talking in terms of modes presentation, but even the leeriest acknowledge that it is unproblematic with the right caveats—(e.g.) that the relevant mode of presentation isn't equivalent to a description, or that it needn't refer to something other than *itself* (cf. Levin, 2006, e.g. note 83).

against it. But disvaluing an experience-type might just as easily turn out not to have that entailment, and it's not obvious how to determine which verdict is right. So I'll now turn directly to the question of what it takes to represent someone else's token *instance* of a given experience-type with full understanding.

Note, first, that the central lemma of this paper—that full understanding is compatible only with coming to *dis*value someone's suffering intrinsically—just concerns (i) someone other than ourselves (ii) experiencing an instance of *suffering*. In other words (as I noted above), we are abstracting away from who the sufferer specifically is, and even from how their specific instance of suffering differs from other instances of suffering—so long as (following §1.1) we represent their experience as having some specific, non-zero degree of overall emotional aversiveness. My question just concerns what evaluative attitudes we can form with respect to the bare fact that someone other than ourselves is suffering, and who the sufferer specifically is doesn't have any direct bearing on that question.[30]

It's worth noting, in particular, that understanding the bare fact that they are the sufferer is very unlikely to require an experiential mode of presentation for the *person* as such.[31] It is controversial, after all, whether specific 'subjects of experience' (our *selves*, in other words) are themselves direct objects of our experience in the first place (cf. "Anattalakkhaṇa Sutta: The Characteristic of Not-Self", 2018; Nāgārjuna, 1934; Hume, 1738, I, IV, §6; Dennett, 1992; see Duncan, 2019 for an argument against such views).

[30] Another way of putting the point is that given the question at issue, all that matters about the specific sufferer is that they are the one suffering. But fully understanding *that* about them doesn't require that we understand—let alone *know* (cf. Benton, 2017)—the sufferer or 'who they are' in any way that involves appreciating their specific differences from other sufferers. Their 'numerical' difference is enough.

[31] In contrast, as I'll be arguing in §2.3, fully understanding the sufferer's *being an experiencer* arguably (i) *is* a necessary condition on fully understanding the fact that someone other than ourselves is suffering and (ii) does seem to require an experiential mode of presentation—though still probably not of the sufferer as such.

In any case, I'll hereafter be assuming that being a minimally competent user of the sufferer's name—which, following Kripke (1980), for instance, likely requires little more than that we be properly situated in a causal chain of name-users leading back to the individual's 'baptism' with that name—is (more than) enough. Fully understanding the expression "someone who isn't me" would do just as well.

The main things we need to understand, then, are just the experience-type *suffering* and the fact that the relevant sufferer's experience is an instance of that type. My claim in what follows is that fully understanding someone else's experience as an instance of a given experience-type requires an experiential mode of presentation for their instance itself: it requires a *token* phenomenal concept of their experience.

At the outset, it might seem to be enough, for full understanding, just to represent the *type* experientially. Suppose, for instance, that the state of affairs we're trying to understand is the fact that the long-suffering Job[32] is being subjected to an instance of the mealy, unpleasantly soft mouth-feel of an unexpectedly over-ripe Red Delicious apple. It might seem to be enough to recall, vividly, an instance of that kind of experience and think,

‹Job is having an experience of ⟨this type⟩›.

The representational target, again, is *Job's experience's being an instance of mealy apple mouth-feel*. So we might think of our representation as having two argument places: one for the experiencer (Job), and one for the experience-type of which his experience is an instance (mealy apple mouth-feel).

[32]See *The Book of Job* in any version of the Bible or the Tanakh. Job is famous mostly for suffering a lot, so I use him as the main 'arbitrary person other than ourselves' in the rest of the paper. Any apparent relevance of the details of Biblical Job's suffering or his relationships with his friends (let alone God or the devil) is purely accidental.

The problem is that we might be able to understand fully both the experience-type *mealy apple mouth-feel* and *Job's being the experiencer* but fail to appreciate fully the nature of the *instantiation* of an experiential type. What matters, in particular, is that an experience-type's being instantiated is itself an essentially experiential fact.

The result is that even if we fully understand the content of both argument places—and even *accurately* represent the instantiation relation between them—we will still fail to understand fully the fact as a whole unless we also represent the instantiation relation itself experientially. Otherwise, with respect to the experience-type's actually being (experientially) instantiated, we are at best like the version of Mary, above, who maybe *could* summon up a vivid memory of blueness on command but wasn't currently doing so.

Full understanding, then, requires that we represent the given experience-instantiation itself experientially. But to do that—to represent experientially, say, Job's instantiating the experience-type *mealy apple mouth-feel*—just is to represent his experience with a token phenomenal concept.

It turns out, then, that there is a version of the thought ‹Job is having an experience of ɬthis typeþ› that *can* suffice for full understanding: namely, a version in which we deploy a phenomenal concept for the *instantiation* of the relevant type, too. And that makes that version of the thought equivalent to ‹Job is having ɬhis experience of this typeþ› or ‹Job is having ɬthisþ›, where ‹ɬthisþ› is a token phenomenal concept of his instance of mealy apple mouth-feel—or even, if you'll allow thoughts with less obviously propositional structure—just ‹ɬJob's instance of mealy apple mouth-feelþ›.[33]

[33]In §1.3, I claimed that adding just a small amount of positive or even neutral affect to a given suffering experience makes it an importantly different suffering experience. So you might wonder why the same shouldn't hold when Bildad adds, to his phenomenal concept of Job's token experience, his mental label of the experience as 'happening in Job': why doesn't that labeling undermine Bildad's ability to maintain an accurate simulation of Job's overall affective state? The worry is particularly clear if Job is having such an all-consuming suffering experience that he momentarily lacks the cognitive bandwidth even to represent

2.3 Two ways to represent someone else's experience

It's worth taking a step back to get a more intuitive grip on what we gain by representing Job's experience in that way.

Again, for our purposes, all that matters about Job is that he is someone other than ourselves. But so far I've been talking as if representing 'someone other than ourselves' as the relevant experiencer were a straightforward matter.

It isn't, it turns out: even leaving out everything specific to Job, fully appreciating that he is the experiencer requires fully appreciating the fact that someone other than ourselves exists, is just as truly an experiencer as ourselves—is just as much a 'seat of experience'—*and* is having an experience of the relevant type. Perhaps despite initial appearances, that turns out to be a highly non-trivial task.[34] More to the point, it's a task that upon closer inspection is itself impossible to accomplish without using token phenomenal concepts.

To see why, consider Job's experiencing an instance of mealy apple mouth-feel, again, with his friends Eliphaz the Temanite and Bildad the Shuhite representing that fact in the two different ways at issue. Suppose, first, that Eliphaz represents Job's experience by summoning up a vivid memory of mealy apple mouth-feel and thinking ‹Job is having an instance

himself as Job. The answer to this question, as I see it, hinges on the affective contribution of the 'labeling' at issue. First, to the degree that our knowledge of our own identity (as (e.g.) *Job*, that is, rather than just as "me"; cf. note 36) can be wholly *implicit*, *non-experiential*, or just *non-affective* (rather than, say, merely affectively neutral), to that degree there should be no issue: I see no reason to think Bildad's 'representing this as Job's' couldn't be just as minimal as Job's own self-representation. But suppose Job's experience is, *necessarily*, one that would be affectively changed by 'self-labeling' of even the most minimal sort. My view entails that in that case, Bildad would indeed be barred from simultaneously fully accurately simulating Job's suffering experience and knowingly using that simulation as a token phenomenal concept of *Job*'s suffering. Note, though, that even if that were right and Bildad's experiential representation had to 'underplay' Job's suffering by the relevant amount, in such cases, the argument in §1.3 would still entail that he could only come to intrinsically disvalue Job's suffering, *so* represented. The affective contributions of valuing and evaluative neutrality would still make it impossible for Bildad to sustain that best *possible* representation of Job's suffering. Note, finally, that despite not embodying full understanding, that best possible representation could also be unproblematic in terms of *accuracy*: Bildad could think of Job's experience as (say) ‹even worse than ⟨this⟩›. Thanks to an anonymous reviewer for bringing these very interesting issues to my attention.

[34] My comments here are closely related to the so-called 'conceptual problem of other minds' (e.g. Avramides, 2020, §2). Thanks to Haigen Messerian for pointing out the relevance of this literature.

of ⟨this type⟩. Suppose Eliphaz's experience of thinking that thought—other than the component corresponding to his type phenomenal concept—is just the experience of subvocalized versions of the relevant words. He is conceptually competent with the words he is subvocalizing, but their accompanying modes of presentation aren't relevantly experiential.

Bildad, suppose, also has a vivid memory or simulation of the mouth-feel of a mealy apple. But in him it plays the role of a *token* phenomenal concept of Job's experience: he represents Job's instance of mealy apple mouth-feel itself with an experiential mode of presentation. Bildad's representation is thus equivalent to ‹⟨Job's instance of mealy apple mouth-feel⟩›.

The first thing to note is that Eliphaz's representation is relatively abstract. We can suppose he understands the experience-type and (even arbitrarily specific) phenomenal quality of Job's experience perfectly vividly, but his representation of Job's actually *experiencing* an instance of that type is verbally mediated, centrally featuring an explicit deployment the 'lexical' concept ‹instance›. The key result is that, *experientially*, Eliphaz's thought is akin to ‹⟨this experience-type⟩ has some feature or another›.

The problem is that failing to represent the experience-type's instantiation experientially means that Eliphaz's thought is compatible with having only the dimmest understanding that other people exist and have experiences. So, if this thought, experienced in the way described, is the only way Eliphaz represents Job's situation, then he fails to appreciate fully that Job is just as much a 'seat of experience' as he (Eliphaz) is.[35]

[35] Again, my talk of understanding people as 'seats of experience' doesn't entail that subjects of experience or selves are experienced directly, as *objects of experience*, rather than being more like (say) 'locations' in which a person's experiences happen (cf. §2.2). It is their *actually happening, in a person distinct from ourselves*, that Eliphaz is failing to appreciate fully.

Bildad's representation, in contrast, is like the ones we might generate when reading an engrossing biography: as we read about the person's experiences, many of us spontaneously internally simulate those experiences, as a part of our representation of *their experiencing what they experienced*: we are implicitly deploying an experiential mode of presentation—a token phenomenal concept, in other words—to represent those actual events (cf. Walton, 2015).

The important result is that token phenomenal concepts allow Bildad to represent Job's actually experiencing an instance of the relevant experience-type with more complete understanding, insofar as his representation is *experientially explicit* about this essentially experiential relation. So Bildad's token phenomenal concept puts him in a position to attribute to Job's token experience, in a 'cognitively direct' way and while fully experientially understanding it, the experiential feature of *instantiating* mealy apple mouth-feel.

Now, that doesn't guarantee that Bildad fully appreciates the fact that Job is just as truly an experiencer as Bildad is. Experientially, Bildad's thought might be equivalent to ‹ðthis experience-type's being instantiated in its experiencerþ has some feature or another›, where ‹has some feature or another› corresponds to Bildad's representation of the fact that the experiencer is Job. So it's possible, even with an experiential mode of presentation for the instantiation of an experience-type, not to be particularly precise about who is having (or had) the relevant token experiences in the first place—let alone being specifically clear on the fact that we and the target experiencer both exist and are distinct individuals.[36]

[36]See Kripke (1982, postscript), Nussbaum (2001, 327-328), and Wittgenstein (1953, especially §§293-317) for more on this and related issues (and cf. note 34). Kripke's reading of Wittgenstein in particular draws out the difficulty of identifying even ourselves: while we know 'I am myself' (the subject of 'ðthese experiencesþ'), we might fail to know virtually everything else about ourselves—including "whether [we are, say,] Jones or the person in the corner" (Kripke, 1982, 144). Thanks to Ken Walton for very helpful conversation on these and related issues.

That said, Bildad gets much closer to full understanding than Eliphaz, all else being equal. By vividly representing the experiential nature of Job's instantiation of mealy apple mouth-feel, Bildad understands, experientially, that a mealy apple mouth-feel experience is *happening in an experiencer*, whereas Eliphaz's thought only experientially represents the experience-type in the abstract.

2.4 Normalizing other-regarding token phenomenal concepts

In short, then, we need token phenomenal concepts for effectively the same reason we need type phenomenal concepts: namely, to appreciate fully the experiential nature of their referents.

Even so, it's natural to be especially skeptical about token phenomenal concepts of other people's experiences—maybe because it can be hard to imagine what it's like to deploy them, which might then fuel skepticism about their coherence in the first place.

The worry, first of all, seems to stem from the fact that other-regarding token phenomenal concepts represent something that is experiential (say, Job's suffering) but which is distinct from the experience *constituting the phenomenal concept itself* (say, Bildad's simulated or remembered token suffering experience).

A first response is thus to note that type phenomenal concepts have that same feature: they represent something—an experience-*type*—that is distinct from the *token* experience out of which which those concepts are formed. But there is also direct precedent for positing *token* phenomenal concepts that designate an experience (albeit still one happening in the representer of the experience) distinct from the experience constituting the token phenomenal concept itself. Balog, for instance, might even think *all* token phenomenal concepts work that way (see Balog, 2012, 36-37).

Second, note that we seem to do something closely analogous to deploying other-regarding token phenomenal concepts whenever we remember our own past experiences or anticipate or ponder future experiences (or even merely possible ones) using experiential modes of presentation.

Think back to the last time you had a burrito at your favorite taquería, for instance (or any other specific food on any specific day). It's not uncommon for vivid visual, olfactory, and gustatory 'images'[37] to be a part of such memories. And most of the time, I submit, we have no separate thought like ‹THAT BURRITO EXPERIENCE WAS OF ⊰THIS TYPE⊱›; the images are just part of our mode of presentation of the token experience we had at the taquería (cf. Fernández, 2017; Montero, 2020).

In memories, then, the content of the *time of occurrence* argument-place in our representation is the time we're remembering, while the time of occurrence of the memory itself is (say) *now*. That makes memory strongly parallel to token phenomenal concepts of other people's experiences. The difference is just that in the latter, it is the content of the *experiencer* argument place, rather than the *time* argument place, that differs from the corresponding feature of the representation itself. The experiencer of Job's suffering is Job, and the experiencer of the 'representational' remembered or simulated suffering—the one constituting Bildad's phenomenal concept of Job's suffering—is Bildad.

In both cases, then, a property of the relevant *represented* experience is importantly different from the corresponding property of the experience constituting the *experiential representation* itself. So if (as we seem to take for granted) there is nothing suspect about that feature of normal memory

[37]Barring aphantasia, that is. Aphantasics (see (e.g.) Whiteley, 2020) report not experiencing *phenomenological* mental imagery. They might nonetheless still meet my conditions for full understanding if (i) they are actively having an experience of the target type (e.g., by having a mealy apple in their mouth) or if (ii) there are forms of mental imagery that are first-personal or experiential despite still being 'phenomenologically blank.'

and prospective and counterfactual imagination, then there also shouldn't be anything suspect about token phenomenal concepts of other people's experiences, unfamiliar though they may seem.

Third and finally, though, other-regarding token phenomenal concepts needn't even seem so unfamiliar: I suspect most of us already use them quite often in the context (similar to the 'gripping biography' context I discussed earlier) of hearing a friend describe an experience they went through.[38]

3 Caligula and the (normative?) 'Mary's room' of suffering

We now have my two central premises in place. First, if we form an evaluative attitude towards it at all, it's impossible not to be intrinsically against any suffering we represent experientially *as suffering*—as exemplified by the case of our own, occurrent suffering (§1): otherwise, given the relevant 'test' conditions for intrinsic valuing, the attitude we end up forming can only be towards a less overall emotionally aversive experience than the suffering experience at issue. Second, fully understanding the fact that, say, Job is suffering requires that we represent his token instance of suffering experientially, by means of a token phenomenal concept (§2). It follows that anyone—sadist or saint—who fully understands the fact that someone else is suffering will care: they will necessarily be disposed to be intrinsically against the other person's suffering.

[38]Suppose a friend tells us that a stranger on a scooter zipped by dangerously close to them on the sidewalk yesterday afternoon and caused them to trip sideways into a balloon-animal vendor's massive bouquet of balloon-animals, and from there into a deep patch of wet cement. It wouldn't be unusual for their story to evoke a visual-auditory-vestibular-tactile-emotional simulation in us—and for that to constitute, at least partly, our representation of their specific experience. Again, we needn't think of what we're doing as representing only the *type* their experience instantiated, such that we then add to ourselves, in some separate thought-component, that our friend had an instance of ‹┤THAT TYPE├›. We can coherently take our spontaneous simulation itself to represent their specific experience, with the visual-auditory-vestibular-tactile-emotional imagery their story evokes constituting a phenomenal concept equivalent to ‹┤THEIR TOKEN EXPERIENCE├›.

3.1 Affect-components and impossible understanding, revisited

That might sound very counterintuitive, given that we're no longer talking about our own suffering. Suppose Caligula fully understands the fact that Job is suffering, using a token phenomenal concept constituted by a specific degree of overall emotional aversiveness. Why should coming to value (positively or neutrally) what he thereby represents necessarily lessen that emotional aversiveness? It's natural to think that the representation of Job's suffering could be 'affectively encapsulated,' in other words, such that the rest of Caligula's affective experience—including the affective contribution of his evaluative attitudes under the relevant 'test' conditions—would be free to vary.

My response is that, first, as long as two affective experience-components are happening within one genuinely unified mind and mental state, they will necessarily contribute to that mental state's overall balance of positive, negative, and neutral affect. But, in keeping with §1.3, if Caligula is to meet the conditions for intrinsically valuing Job's suffering, then he must be able to feel, *simultaneously* and *within a single, unified mental state*, both the positive affective contribution of valuing and the affective contribution of what would *on its own* be an accurate simulation of Job's suffering experience, with its specific degree of overall emotional aversiveness. Otherwise, it isn't Caligula as such but at most an encapsulated component of Caligula's mind (opaque to Caligula himself) that fully understands.[39] So the degree of overall emotional aversiveness of *Caligula's* experience—and thus of the token phenomenal concepts of overall affective tone that he can form—will necessarily be less than Job's.

[39]See Atiq's 'split-brain Mary' (Atiq, 2021, §3) for an evocative related case.

A last thing to note is that while we can probably get close, especially with the help of sincere descriptions from the sufferer, it's very unlikely that we or anyone can—*knowingly*—simulate or recall an instance of exactly the same degree of overall emotional aversiveness instantiated by someone else's suffering experience. It is nonetheless in a clear sense 'humanly possible' to do so, even if we can't know we're pulling it off. More to the point, though, even if we're certainly getting it wrong, intrinsic disvaluing is still the only intrinsic evaluative attitude we can form towards their suffering, *as we represent it*, if we represent it experientially. All other attitudes would, in the relevant 'test' conditions, force us to (at best) 'change the subject' on ourselves, by changing the overall degree of emotional aversiveness experientially available for forming our phenomenal concept of the suffering we were supposed to be considering. So the fullest possible human understanding—but also *any* understanding constituted by a token phenomenal concept representing a specific degree of overall emotional aversiveness (accurately or not)—makes it impossible to form any intrinsic evaluative attitude towards someone else's suffering other than being intrinsically against it.[40]

3.2 Establishing and explaining suffering's badness

With that in place, all we need for the conclusion that suffering is bad is the following:

[40]To be clear, there is no guarantee that anyone will actually come to disvalue someone else's suffering, despite this conclusion, because there is no guarantee that anyone will come to understand the fact that someone else is suffering in the ways at issue. It is relatively easy to remember that we ourselves don't enjoy suffering, that it 'feels bad,' and that other people probably feel the same about their own suffering—all without using token phenomenal concepts. But those less demanding forms of understanding are in principle quite compatible with valuing other people's (or even our own) suffering.

It's also worth noting that my argument didn't rely on any claims about Caligula's *identifying with* or valuing the *person* who is suffering (Job, say), or about anyone's coming to *believe* that suffering is bad—whether prudentially bad for the sufferer, morally bad for us to cause or ignore, or bad in any other way. As far as my argument is concerned, full understanding could be compatible with Caligula's actively *disidentifying* from the sufferer, explicitly disvaluing Job as a person, and believing his suffering is objectively *good*. All three are probably unlikely, if he does come to disvalue Job's suffering; it all depends on the exact relationship between disvaluing someone's suffering and believing it's bad, valuing the sufferer, and 'identifying' with the sufferer. Thanks to an anonymous reviewer for inviting me to clarify these points.

> If some (non-'centered'[41]) state of affairs is such that necessarily, everyone who fully understood it would be disposed to disvalue it intrinsically, then that state of affairs is objectively, agent-neutrally bad.

A broad swath of theorists already endorses something like this premise. Notably, denying that anything could meet its objectivity-related 'necessary, universal valuing' condition seems to be what makes some people error theorists (e.g. Joyce, 2001; Mackie, 1977; probably also Wittgenstein, 1965), others various kinds of anti-realist (cf. Street, 2016), and arguably others quasi-realist expressivists (Blackburn, 1984; Gibbard, 2002, e.g. 54).

If the arguments in the previous sections were sound, then such philosophers might happily embrace a very minimal version of moral realism—at least on the assumption that the objective, agent-neutral badness of anyone's suffering would constitute a specifically *moral* fact (cf. Machery, 2012; Marshall, 2018, e.g., ch. 9). My main goal in this paper was to give an argument that might convince global moral skeptics,[42] so since many or most of them accept the above sufficient condition for objective, agent-neutral badness, I may now have done all I could on that score.[43]

---

[41] Importantly, something's being *agent-neutrally* (dis)valuable—roughly, being (dis)valuable such that the normative reasons entailed for a given agent are "[not] a function of the features of that agent as such" (Ridge, 2023, §5)—requires that the candidate (dis)values not include anything 'agent-relative' or 'centered' (Egan & Titelbaum, 2022, §3.1). So the 'centered property' (Lewis, 1979) of *suffering*—the property such that to disvalue it is to disvalue *(ourselves) being subjected to suffering*—is not a candidate agent-neutral disvalue, but *Job*'s suffering is.

[42] My goal thus differs significantly from that of Andrew Lee (in press) and Thomas Nagel; theirs is "to get rid of the obstacles to the admission of the obvious" (Nagel, 1980, 109, as quoted in Lee, forthcoming; cf. Goldstein, 1989; Nagel, 1986). Something similar is arguably also true for Guy Kahane (2010).

[43] Others—especially non-naturalists (e.g. Enoch, 2017)—might not have much attraction to the idea that something's being objectively valuable is guaranteed whenever necessarily everyone who understood it would be disposed to value it intrinsically. In particular, anyone leery of the mere whiff of a suggestion that people's value-dispositions are what *explain* or *ground* facts about objective goodness or badness might be turned off. So, to be clear, my goal has not been to say what explains suffering's badness, in this sense of 'explains'; my argument is thus compatible with any explanation for *why* meeting that condition would guarantee suffering's badness. Note also that many non-naturalists will have very good reason not to reject the sufficient condition at issue (minus offending 'metaphysical whiff'). Many non-naturalists give pride of place to everyday spontaneous evaluative attitudes and intuitions, corrected by reflective equilibrium—no doubt with the evidential value of a given attitude scaled by how unbiased and clear-sighted it is. So it's hard to imagine any attitude evidentially better off than one necessarily shared by everyone who fully understood what they were considering and formed an evaluative attitude about it at all.

That said, my arguments in this paper fit especially well with one appealing story about what *explains* suffering's badness.[44] First, §1's take on suffering and valuing aligns nicely with the idea that the 'reason-giving force' of suffering is grounded in the fact that it *feels really bad* to the one suffering. It's natural to think, in particular, that suffering's felt badness not only makes it reasonable for the sufferer to want it to stop but also grounds all further normative facts about suffering (cf. Kahane, 2009).[45]

If that's right, then my paper can be seen as pointing to a partial explanation for one derivative (if important) class of further normative facts: in particular, it might help explain why the reason-giving force of anyone's suffering also 'applies' to or generates recognizably practical-rationality-related reasons for other people. The thought is that if fully understanding the fact that someone else is suffering guarantees that we will be disposed to disvalue it, then coming to value it positively or neutrally entails that we don't even fully 'know what we're talking about' (cf. Sobel, 2009; Lewis, 1989) when considering their suffering. If Caligula values Job's suffering, it's not actually *as* 'Job's instance of suffering' (as such) that Caligula values it, but at most under some importantly inaccurate or incomplete mode of presentation.

We're arguably all committed to holding our attitudes—attitudes that we ourselves take to be about something *as it in fact is*[46]—accountable to the consequences of fully understanding what we're evaluating in the first place. (Think of someone pontificating on the quality of a film they haven't seen or the taste of a food they've never tried or only dimly remember.)

[44] See Marshall (2018, especially ch. 10) for an extended treatment of related questions.

[45] A recent view related to Kahane's is that, *contra* Moore (1903) and others, negative affective states are intrinsically bad *by definition* (Rawlette, 2020).

[46] We might *not* take our attitude to be about someone else's suffering as such, in a given case. But note that we presumably always implicitly take our attitudes to be about *something* 'as such' (e.g. about Job's suffering's *being deserved* or its *allowing me to take his favorite seat at the feast*), and the question of this paper is whether anything—and suffering in particular—is such that understanding *it* ('for what it is') guarantees universal (dis)valuing.

And if so, then there's a fairly straightforward sense in which we would be going against our own, 'non-alien' (cf. Railton, 1986) reasons if, faced with someone else's suffering, we failed to respond to its reason-giving force by failing to be intrinsically against it.[47]

In short, requirements like the one to 'know what we're talking about' might be understood as establishing a kind of normative channel, through which the normative force of more ground-floor (say, *experiential*) values are passed on to other agents. This is a very speculative proposal, though, and it might not be possible to identify any candidate values other than suffering that would fit the specific pattern at issue—other than, for strictly parallel reasons, happiness.[48]

Conclusion

In conclusion, suffering is bad. More radically, but in support of that pretheoretically uncontroversial conclusion: when it comes to other people's suffering, it's impossible to come to any evaluative attitude other than being intrinsically against it, as long as we fully understand the fact that they are suffering.

My argument has its share of potentially controversial premises. Probably the most in need of further support is my claim that fully understanding something essentially experiential requires not just that we *can* vividly

[47]I suspect this section's observations might also secure a recognizable *sense* in which something that met that condition would be objectively bad, insofar as it might secure a way in which we would be *rationally criticizable* for not disvaluing it. But that kind of badness might be somewhat less forcefully reason-giving than 'full-on, morally demanding badness.' The reason is that on their own, universalizing-, interpersonal consistency-, or 'full understanding'-related rational requirements don't seem to be in a position to explain why suffering is more important than, say, chocolate. A requirement to 'know what we're talking about,' for instance, makes no distinction between Caligula's intrinsically valuing his victim's suffering and a child's being intrinsically against the taste of chocolate before ever tasting it. Both valuers don't fully know what they're talking about, and potentially to the same degree. So the main normative work, in terms of explaining the difference in normative 'weight' between such cases, would seemingly have to be done not by epistemic principles but by intrinsic features of the specific values and disvalues at issue. The reason suffering is more important than chocolate, for instance, might come down to the difference in their *felt* importance, from the inside.

[48]Happiness, I strongly suspect, is good.

remember or simulate the relevant experiential qualities, but that we actually be doing so, right in the moment of understanding. Others have defended that premise in greater depth than I did in §2.1, as I noted, but my conclusion should in any case be understood as conditional on the success of those and other arguments.[49]

My argument also left many questions relevant to suffering's badness unanswered, including whether what I've said is compatible with retributivist justifications for punishment,[50] what follows for what we're morally required to *do* about someone's suffering, whether the argument extends to non-human suffering (though I see no reason why it wouldn't), and in general most of the relevant normative ethical questions worth asking.

In sum, though, I hope I've motivated the idea that, when we consider someone else's suffering, intrinsically disvaluing it is not just a *likely* or *morally fitting* reaction but the only evaluative attitude towards their suffering (as such) that we can form at all—that is, if we fully understand that they are suffering in the first place. Any other attitude, I've argued, would make it impossible to continue to represent their specific suffering experience accurately, with its specific degree of overall emotional aversiveness.

Failing that, I hope I've at least highlighted the difficulty—and perhaps the glimmer of the possibility of overcoming the difficulty—of establishing and explaining this seemingly obvious fact that suffering is bad.

Acknowledgments This paper is the product of an unusually large amount of invaluable feedback. For their written comments, special thanks to Jamie Dreier,

[49]Note that even if they all fail, it seems overwhelmingly likely that 'actively experiential' representation at least shouldn't *hurt*, epistemically (all else being equal). And it would remain an interesting—if somewhat harder to interpret—result, if anyone who understood in this particular way that is *at least* as good as all epistemic alternatives would be disposed to intrinsically disvalue everyone's suffering.

[50]My paper raises two different *prima facie* challenges for retributivism. First, If endorsing retributivism entails intrinsically valuing (*qua* suffering) the suffering of wrongdoers, then my argument might entail that it is impossible to endorse retributivism while fully 'knowing what we're talking about.' And second, if people only cause suffering to others (in the way that would make them deserve punishment) to the degree that they fail to disvalue their victims' suffering (perhaps specifically: *relative to their own*), then my argument might entail that no one ever fully knowingly causes suffering in the way that makes them deserve punishment.

Chris Hill, Bernard Reginster, Michael Frank, and Peter Railton, as well as (in roughly chronological order of feedback) David Christensen, Alastair Norcross, Griffin Klemick, Ken Walton, Sarah Buss, Colin Marshall, Matt Duncan, Peter van Elswyk, Justin Dealy, Dan Burnston, and anonymous reviewers at Synthese and other journals. Thanks also to members of the audiences at Brown University's Philosophy Graduate Forum, Brown's Laboratory for Neural Computation and Cognition, the University of Michigan at Ann Arbor's Graduate Student Working Group, the 2016 Objectivity in Ethics conference hosted at the University of Utrecht, the 2018 Rocky Mountain Ethics Congress (especially Andrew Sepielli), and the University of Toronto's 2018 PsyPhi graduate conference. Finally, for their many in-depth conversations with me about this project (and about suffering more generally), thanks to Mark Ho, Yongming Han, Taylor White, Harrison Ritz, Chad Marxen, Kevin Blackwell, Emma Hardy, Nick Alvarez, Ruth Kim, Orfeas Zormpalas, Brett Topey, Jonathan Courtney, Mary Renaud, and Micaela Quintana.

Conflicts of interest declaration

The author has no relevant financial or non-financial interests to disclose.

References

Anattalakkhaṇa sutta: The characteristic of not-self (S. Bhikkhu, Trans.). (2018). *Saṁyutta-nikāya*, *22.59*. Retrieved from https://suttacentral.net/sn22.59/en/sujato

Arpaly, N., & Schroeder, T. (2013). *In praise of desire*. Oxford University Press.

Atiq, E. (2021). Acquaintance, knowledge, and value. *Synthese*, *199*(5-6), 14035–14062.

Atiq, E., & Duncan, M. (in press). I feel your pain: Acquaintance & the limits of empathy. *Oxford Studies in Philosophy of Mind*, *4*.

Atisha (2017). *The lamp for the path to enlightenment* (K.D.K. Choephel, Trans.). Dharma Ebooks. (Commentary by Jamgön Kongtrul Lodrö Thaye)

Avramides, A. (2020). Other Minds. E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2020 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/other-minds/.

Bailey, O. (2022). Empathy and the value of humane understanding. *Philosophy and Phenomenological Research*, *104*(1), 50–65.

Bain, D. (2017). Evaluativist accounts of pain's unpleasantness. J. Corns (Ed.), *The Routledge handbook of the philosophy of pain* (pp. 40–50). Routledge.

Bain, D., Brady, M.S., & Corns, J. (Eds.). (2020). *Philosophy of suffering*. Routledge.

Balog, K. (2009). Phenomenal concepts. B. McLaughlin, A. Beckermann, & S. Walter (Eds.), *The Oxford handbook of philosophy of mind* (p. 298-312). Oxford University Press.

Balog, K. (2012). Acquaintance and the mind-body problem. S. Gozzano & C.S. Hill (Eds.), *New perspectives on type identity: The mental and the physical* (pp. 16–43). Cambridge University Press.

Barlassina, L., & Gordon, R.M. (2017). Folk Psychology as Mental Simulation. E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2017 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/folkpsych-simulation/.

Benton, M.A. (2017). Epistemology personalized. *Philosophical Quarterly*, *67*(269), 813–834.

Blackburn, S. (1984). *Spreading the word: Groundings in the philosophy of language*. Clarendon Press.

Carruthers, P. (2004). Suffering without subjectivity. *Philosophical Studies*, *121*(2), 99–125.

Chalmers, D. (2003). The content and epistemology of phenomenal belief. Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 220–72). Oxford University Press.

Clark, A. (2000). *A theory of sentience*. Clarendon press.

Clark, A. (2005). Painfulness is not a quale. M. Aydede (Ed.), *Pain: New essays on its nature and the methodology of its study.* Bradford Book/Mit Press.

Coleman, S. (2019). Painfulness, suffering, and consciousness. D. Bain, J. Corns, & M. Brady (Eds.), *Philosophy of suffering* (pp. 55–74).

Corns, J. (2021). Suffering as significantly disrupted agency. *Philosophy and Phenomenological Research*.

Dennett, D.C. (1992). The self as a center of narrative gravity. F.S. Kessel, P.M. Cole, & D.L. Johnson (Eds.), *Self and consciousness: Multiple perspectives* (pp. 103–115). Lawrence Erlbaum.

Duncan, M. (2019). The self shows up in experience. *Review of Philosophy and Psychology*, *10*(2), 299–318.

Duncan, M. (2020). Knowledge of things. *Synthese*, *197*(8), 3559–3592.

Egan, A., & Titelbaum, M.G. (2022). Self-Locating Beliefs. E.N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Winter 2022 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2022/entries/self-locating-beliefs/.

Elgin, C.Z. (1983). *With reference to reference*. Hackett Publishing Company.

Enoch, D. (2017). Non-naturalistic realism in metaethics. T. McPherson & D. Plunkett (Eds.), *The routledge handbook of metaethics* (pp. 29–42). Routledge.

Fernández, J. (2017). The intentional objects of memory. S.B..K. Michaelian (Ed.), *The Routledge handbook of philosophy of memory* (pp. 88–100).

Francescotti, R. (1999). Mere cambridge properties. *American Philosophical Quarterly*, *36*(4), 295–308.

Gibbard, A. (1999). Morality as consistency in living. *Ethics*, *110*(1), 140–164.

Gibbard, A. (2002). The reasons of a living being. *Proceedings and Addresses of the American Philosophical Association*, *76*(2), 49–60.

Goldstein, I. (1989). Pleasure and pain: Unconditional intrinsic values. *Philosophy and Phenomenological Research*, *50*(December), 255–276.

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. Hackett Publishing Company.

Grahek, N. (2007). *Feeling pain and being in pain*. MIT Press.

Grant, J.A. (2014). Meditative analgesia. *Annals of the New York Academy of Sciences*, *1307*(1), 55–63.

Grimm, S. (2021). Understanding. E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/understanding/.

Hills, A. (2016). Understanding why. *Noûs*, *50*(4), 661–688.

Hume, D. (1738). *A treatise of human nature*. Oxford University Press.

Jackson, F. (1986). What mary didn't know. *Journal of Philosophy*, *83*(May), 291–295.

Joyce, R. (2001). *The myth of morality*. Cambridge University Press.

Kahane, G. (2009). Pain, dislike and experience. *Utilitas*, *21*(3), 327–336.

Kahane, G. (2010). Feeling pain for the very first time: The normative knowledge argument. *Philosophy and Phenomenological Research*, *80*(1), 20–49.

Kant, I. (2002). *Groundwork for the metaphysics of morals*. Oxford University Press.

Kauppinen, A. (2020). The world according to suffering. M.S. Brady, D. Bain, & J. Corns (Eds.), *The philosophy of suffering*. Routledge.

Klein, C. (2015). *What the body commands*. MIT Press.

Korsgaard, C.M. (1996). *The sources of normativity*. Cambridge University Press.

Kripke, S.A. (1980). *Naming and necessity* (D. Byrne & M. Kölbel, Eds.). Cambridge, MA: Harvard University Press.

Kripke, S.A. (1982). *Wittgenstein on rules and private language*. Harvard University Press.

Lawler, I. (2018). Understanding why, knowing why, and cognitive achievements. *Synthese*(11), 4583–4603.

Lee, A.Y. (in press). Metaethical experientialism. G. Lee & A. Pautz (Eds.), *The importance of being conscious.* Oxford University Press.

Levin, J. (2006). What is a phenomenal concept? T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism.* Oxford University Press.

Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review*, *88*(4), 513–543.

Lewis, D. (1989). Dispositional theories of value. *Proceedings of the Aristotelian Society*, *63*, 89–174.

Lewis, D. (1990). What experience teaches. W.G. Lycan (Ed.), *Mind and cognition* (pp. 499–519). Blackwell.

Loar, B. (1997). Phenomenal states ii. N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates.* MIT Press.

Lyre, H. (2022). Neurophenomenal structuralism: A philosophical agenda for a structuralist neuroscience of consciousness. *Neuroscience of Consciousness*(1).

Machery, E. (2012). Delineating the moral domain. *The Baltic International Yearbook of Cognition, Logic and Communication*, *7*.

Mackie, J.L. (1977). *Ethics*. Penguin.

Marshall, C. (2018). *Compassionate moral realism*. Oxford University Press.

Mayerfeld, J. (1999). *Suffering and moral responsibility*. Oxford University Press.

McRae, E. (2018). Suffering and the six perfections: Using adversity to attain wisdom in mahāyāna buddhist ethics. *Journal of Value Inquiry*, *52*(4), 395–410.

Montero, B.G. (2020). What experience doesn't teach: Pain amnesia and a new paradigm for memory research. *Journal of Consciousness Studies*, *27*(11-12), 102–125.

Moore, G.E. (1903). *Principia ethica* (T. Baldwin, Ed.). Mineola, N.Y.: Dover Publications.

Nāgārjuna (1934). The ratnāvalī of nāgārjuna (G. Tucci, Trans.). *Journal of the Royal Asiatic Society of Great Britain and Ireland*, *66*, 307–325.

Nagel, T. (1970). *The possibility of altruism*. Oxford Clarendon Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*(October), 435–50.

Nagel, T. (1980). The limits of objectivity. *The Tanner lectures on human values*, *1*, 75–139.

Nagel, T. (1986). *The view from nowhere*. Oxford University Press.

Nida-Rümelin, M., & O'Conaill, D. (2021). Qualia: The Knowledge Argument. E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/qualia-knowledge/.

Nietzsche, F. (1887). *On the genealogy of morals: A polemic*. Oxford University Press.

Nussbaum, M.C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge University Press.

Pritchard, D. (2009). Knowledge, understanding and epistemic value. *Royal Institute of Philosophy Supplement*, *64*, 19–43.

Rachels, S. (2003). A defense of two optimistic claims in ethical theory. *Philosophical Studies*, *112*(1), 1–30.

Railton, P. (1986). Facts and values. *Philosophical Topics*, *14*(2), 5–31.

Railton, P. (1999). Normative force and normative freedom. *Ratio*, *12*(4), 320–353.

Rawlette, S.H. (2020). *The feeling of value: Moral realism grounded in phenomenal consciousness*. Self-published.

Ridge, M. (2023). Reasons for Action: Agent-Neutral vs. Agent-Relative. E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2023 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/reasons-agent/.

Rosati, C.S. (2016). Moral Motivation. E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/moral-motivation/.

Sallasutta: An arrow (S. Bhikkhu, Trans.). (2018). *Saṁyuttanikāya*, *36.6*. Retrieved from https://suttacentral.net/sn36.6/en/sujato

Scarantino, A. (2014). The motivational theory of emotions. D. Jacobson & J. D'Arms (Eds.), *Moral psychology and human agency* (pp. 156–185). Oxford University Press.

Schopenhauer, A. (2010). *The world as will and representation* (Vol. 1; J. Norman, A. Welchman, & C. Janaway, Trans.). Cambridge University Press.

Sinhababu, N. (2022). Pleasure is goodness; morality is universal. *Ethical Theory and Moral Practice*, 1–17.

Sliwa, P. (2015). Understanding and knowing. *Proceedings of the Aristotelian Society*, *115*(1pt1), 57–74.

Sliwa, P. (2017). Moral understanding as knowing right from wrong. *Ethics*, *127*(3), 521–552.

Smuts, A. (2007). The paradox of painful art. *Journal of Aesthetic Education*, *41*(3), 59–77.

Sobel, D. (2009). Subjectivism and idealization. *Ethics*, *119*(2), 336–352.

Sosa, E. (2003). Beyond internal foundations to external virtues. L. BonJour & E. Sosa (Eds.), *Epistemic justification: Internalism vs. externalism, foundations vs. virtues.* Wiley-Blackwell.

Street, S. (2016). Constructivism in ethics and the problem of attachment and loss. *Aristotelian Society Supplementary Volume*, *90*(1), 161–189.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Cambridge, Mass.: Harvard University Press.

Stump, E. (2010). *Wandering in darkness: Narrative and the problem of suffering*. Oxford University Press.

Sundström, P. (2011). Phenomenal concepts. *Philosophy Compass*, *6*(4), 267–281.

Walton, K. (2015). Empathy, imagination, and phenomenal concepts. K. Walton (Ed.), *In other shoes: Music, metaphor, empathy, existence* (pp. 1–16). Oxford University Press.

Whiteley, C.M.K. (2020). Aphantasia, imagination and dreaming. *Philosophical Studies*, *178*(6), 1–22.

Whiting, D. (2011). The feeling theory of emotion and the object-directed emotions. *European Journal of Philosophy*, *19*(2), 281–303.

Whiting, D. (2020). *Emotions as original existences: A theory of emotion, motivation and the self*. Palgrave Macmillan.

Wittgenstein, L. (1953). *Philosophical investigations*. Wiley-Blackwell.

Wittgenstein, L. (1965). A lecture on ethics. *Philosophical Review*, *74*(1), 3–12.

Zagzebski, L. (2019). Toward a theory of understanding. *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*, 123–36.