



This is a repository copy of *Conceptual responsibility*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/149152/>

Version: Accepted Version

---

**Article:**

Goetze, T. [orcid.org/0000-0002-3435-3264](https://orcid.org/0000-0002-3435-3264) (2019) Conceptual responsibility. *Inquiry*. ISSN 0020-174X

<https://doi.org/10.1080/0020174X.2019.1658629>

---

This is an Accepted Manuscript of an article published by Taylor & Francis in *Inquiry* on 23rd August 2019, available online:  
<http://www.tandfonline.com/10.1080/0020174X.2019.1658629>

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

**Manuscript ID:** SING-2018-0285.R2

**Title of paper:** Conceptual Responsibility

**Author name:** Trystan S. Goetze

**Author affiliation:** Dalhousie University

**Author email:** [trystan.goetze@dal.ca](mailto:trystan.goetze@dal.ca)

**ORCID:** 0000-0002-3435-3264

## Conceptual Responsibility

### Abstract

Conceptual engineering is concerned with the improvement of our concepts. The motivating thought behind many such projects is that some of our concepts are *defective*. But, if to use a defective concept is to do something wrong, and if to do something wrong one must be in control of what one is doing, there might be no defective concepts, since we typically are not in control of our concept use. To address this problem, this paper turns from appraising the concepts we use to appraising the *people* who use them. First, I outline several ways in which the use of a concept can violate moral standards. Second, I discuss three accounts of moral responsibility, which I call voluntarism, rationalism, and psychologism, arguing that each allows us to find at least some cases where we are responsible for using defective concepts. Third, I answer an objection that because most of our concepts are acquired through processes for which we are not responsible, our use of defective concepts is a matter of bad luck, and not something for which we are responsible after all. Finally, I conclude by discussing some of the ways we may hold people accountable for using defective concepts.

**Word count (body text only):** 8,741

**Word count (body text + notes):** 10,200

**Word count (document total):** 11,209

**Keywords:** conceptual engineering; ameliorative inquiry; moral responsibility; Angela Smith; George Sher

# Conceptual Responsibility<sup>1</sup>

*A smart set of concepts may be a most efficient instrument of corruption.*

Iris Murdoch (2013, 32).

## 1. Introduction: Conceptual Defects and the Point of Conceptual Engineering

Conceptual engineering (or conceptual ethics) is characterised as the ‘critical/constructive enterprise of assessing and improving our representational devices’ (Cappelen 2018, 3). The hope is that we can improve our theories or our societies by improving our concepts. As Sally Haslanger puts it in the course of developing ‘ameliorated’ accounts of the concepts GENDER and RACE, ‘The responsibility is ours to define [our concepts] for our purposes’ (Haslanger 2012, 224).

The idea behind ameliorating our concepts is not merely that some of our concepts work reasonably well for our purposes, but could be made even better. Rather, the motivating thought behind conceptual engineering is that some of our concepts are *defective*. As Herman Cappelen and David Plunkett outline, such defects include: ‘cognitive defects (that undermine our ability

---

1. I am thankful to all those who offered feedback on those earlier drafts of this paper, especially Miranda Fricker, Paul Faulkner, Hallvard Lillehammer, Jennifer Saul, Stephen Laurence, Christopher Bennett, Luca Barlassina, Michael Giudice, an audience at the 2017 Congress of the Canadian Philosophical Association, my fellow students in the postgraduate research seminar at the University of Sheffield from 2014–17, and the attendees of the 2017 Sheffield Philosophy Reading Weekend at the Eyam Youth Hostel. I am also deeply grateful to the editors and an anonymous reviewer for this journal, who made many constructive suggestions that have greatly improved the paper. I also owe my gratitude to the University of Sheffield for funding the first three years of my PhD (when much of the initial research was completed) and for continuing to support my work as an Honorary Research Fellow there, to the Graduate Center of the City University of New York for hosting my study visit in the fall of 2017, to Dalhousie University and the Banting Postdoctoral Fellowships programme for supporting my current position (when this work was completed), and last but not least to my partner, Kat Curwin, for supporting me during my unemployment between these two positions.

to reason properly), moral or political defects (that undermine moral or political values of various sorts), theoretical defects (that undermine progress within some theoretical field), or semantic defects (where the semantic value is incoherent, incomplete, or missing)' (Cappelen and Plunkett, forthcoming 2019, 2 in ms.). So, conceptual engineering is not just engaged in the fine-tuning of abstract machinery; in most cases, the project is an overhaul of, or the construction of a replacement for, a defective concept.

However, this focus on conceptual defects can give rise to an argument that risks undermining the very same motivation for conceptual engineering:<sup>2</sup>

*No Wrong Concepts*

- (1) The claim that motivates conceptual engineering is that some of our concepts are defective, and thus in need of ameliorative work.
- (2) To use a defective concept is to do something wrong.
- (3) But, when the agent  $\phi$ 's, in order for the agent to have done something wrong in  $\phi$ -ing, the agent must have been in control of her  $\phi$ -ing.
- (4) And, our use of particular concepts is not within our control.
- (5) Therefore, no one does anything wrong when they use any concept; and so,
- (6) There can be no defective concepts—and the motivation for conceptual engineering is moot.

There are several ways one could reply to this line of argument. For instance, one could perhaps deny (2), and argue that using a defective concept is to do something *bad* (but not wrong), which one might do without any control over what one is doing. The opponent of conceptual engineering might push back, however, since the distinction between wrongness and badness is not especially clear.<sup>3</sup>

---

2. I am very grateful to an anonymous referee for suggesting this argument as a framing device for the paper.

3. Indeed, some philosophers seem to have the opposite intuition about this distinction. Sher, for example, refers to 'the morally defective acts that render agents blameworthy not as *wrong* acts but rather as *bad* ones' (Sher 2005, 9).

Regardless of the truth of (2) and of the ultimate conclusion of *No Wrong Concepts*, however, the argument also raises a pressing and unexplored problem for conceptual engineering: given that an agent has used a defective concept, how should we appraise *her*? Can we hold her *responsible* for using a defective concept? If (3) and (4) are true, independently of the rest of the argument, then it follows that no one is ever responsible for using defective concepts. And from this, it would seem to follow that we may never *hold* anyone responsible for using defective concepts. For it is a widely-accepted principle that an agent may be held responsible—that is to say, blamed, punished, criticised, or otherwise held *accountable*—only for those things that she is responsible for.

Undermining the possibility that anyone is ever responsible for using a defective concept is less devastating for conceptual engineering than the possibility that there are no defective concepts at all. But conceptual engineering would still be hamstrung in its practical purpose. Part of the practical point of conceptual engineering is to get people to actually use better concepts, but if no one is ever responsible for using defective concepts, then the only permitted approach would be to make exhortations in favour of the ameliorated concepts, while reserving judgement on the people who continue to use defective concepts. We would never be permitted to hold them accountable for doing so—even when the concepts they use are central components of factually mistaken or morally abhorrent views.

In order to defeat *No Wrong Concepts* and to establish the notion of conceptual responsibility, the method I will take in this paper is to argue as follows. First, (4) is false—*sometimes we are in control of the concepts we use*. Second, and more significantly, (3) is false—*control is not a necessary condition for responsibility*. In the background of (3) is a familiar theory of responsibility that is based on the idea that we are only responsible for the things that

we voluntarily bring about. But this kind of account is questionable independently of the upshot for conceptual engineering. The ensuing discussion will show not just that some concepts are indeed defective, but also that we are often responsible for using them, and, thus, may be held accountable for doing so. Third and finally, there are some situations where, even though we are not responsible for something, we may still be held accountable for it, in that we may be asked to *take responsibility for it*.

Here, more precisely, is the plan. In §2, I survey several plausible examples of defective concepts. To focus the discussion of responsibility that follows, I concentrate on cases of *morally* defective concepts. In §§3–5, I discuss several theories of moral responsibility—which I call voluntarism, rationalism, and psychologism—in order to uncover the conditions under which one might be responsible for using a concept. While I argue in favour of the third of these views, each account admits at least some cases where the agent is responsible for using a concept; indeed, the latter two admit a significant range of cases. In §6, I reply to an objection that, regardless of one’s theory of responsibility, the fact that much of one’s conceptual repertoire is acquired through processes that are merely a matter of luck entails that one is never responsible for using a defective concept. In §7, having shown that we can be responsible for using defective concepts, I close by briefly discussing some of the ways in which we might hold people accountable for doing so, whether the concepts’ defects are moral or of some other domain.

## **2. Conceptual Wrongs**

To focus the discussion, it will be helpful to have several cases of defective concepts to return to in the following sections. I will concentrate on ways in which concepts can be *morally* defective in this paper, as it streamlines the discussion of responsibility to follow. (I briefly discuss

cognitively, theoretically, and semantically defective concepts, and whether we might be responsible for using them, towards the end of the paper.)

One way a concept can be morally defective is when the very category it picks out is injurious. Consider, for instance, what Charles Mills says about the concept of a ‘savage’ as applied to indigenous people:

the nonwhite Other is grasped through a historic array of concepts whose common denominator is their subjects’ location on a lower ontological and moral rung...When Thomas Jefferson excoriates the ‘merciless Indian Savages’ in the Declaration of Independence...neither he nor his readers will experience any cognitive dissonance with the earlier claims about the equality of all ‘men’, since savages are not ‘men’ in the full sense. (Mills 2007, 26–7)

The wrong of thinking of an indigenous person using the concept SAVAGE is the implication that she is less than fully human, or otherwise inferior to members of other races.<sup>4</sup> This degradation will likely be expressed in racist discriminatory action on the part of the person using the concept SAVAGE. But it is plausible that *simply thinking* those derogatory thoughts is morally problematic. As Ernesto Garcia remarks, ‘What seems most truly evil about racist action is not so much any particular acts of taunting, racial slurs, or displays of disrespect, but instead...the pernicious underlying attitude: that the racist sincerely believes members of a different race somehow count as less than human’ (Garcia 2002, 202–3). A closeted racist, who thinks of all indigenous people as ‘savages’ but never acts on his disparaging views, would still be doing something morally wrong by using that defective concept.

---

4. We might add that SAVAGE is also semantically defective, since it fails to refer (no one is ‘less than fully human’), theoretically defective, because assuming there is such a category undermines progress in social science, and politically defective, because, as Mills’s passage illustrates, it undermines the political value that all people are equal.



Another way a concept can be morally defective is because of the harms its use ushers in. Of course, not just any harms stemming from the use of a concept would count as morally wrong. Thinking of someone who always shies away from standing up for what is right as a coward, or thinking of a white supremacist as a racist may bring harm to them, for doing so might justify shaming or ostracising them for their actions and views. But such harms would not be *undeserved*. These agents' morally bad character, actions, and beliefs justify the negative evaluation we make of them using the concepts COWARD and RACIST, making the application of these concepts apt, and at least some of the harms they thereby experience permissible.<sup>5</sup> Thus, despite the harm it may produce, using these concepts, in this instance, would not be wrong.

By contrast, consider Talia Mae Bettcher's critique of mainstream sex and gender concepts from the perspective of a trans person. On her account, the prevailing forms of the concepts MALE and FEMALE or MAN and WOMAN are partially defined in terms of the genitals someone has. The content of these concepts thus entails that trans people who have not had sex reassignment surgery do not really have the gender with which they identify and, thus, must be either 'pretending' to have that gender identity, or 'deceiving' themselves or others about their 'true' sex or gender. Bettcher argues that this representation of trans people is harmful in two ways. First, denying that trans people have the gender with which they identify does demonstrable psychological harm, and, moreover, silences trans people's testimony regarding

---

5. I say 'at least some of the harms' and not 'all' the harms, because there will be some threshold at which the harms that come to the coward or the racist because they are so categorised become undeserved. To take an extreme example, it would be impermissible to *murder* them for their cowardice or racism. While it may be true that such extreme harms are produced by the application of these concepts, in the sense that the use of these concepts is upstream in the causal chain leading to the murder, the fact that this harm is quite beyond what is reasonably permitted by identifying someone as a coward or a racist does not, I think, mean that the application of those concepts was wrong after all. Rather, the wrong lies in the murderer taking things too far.

their own experiences: ‘after identity enforcement, nothing we [trans people] might say could possibly matter’ (Bettcher 2007, 51). Second, the notion of trans people as ‘pretenders’ and ‘deceivers’ that follows from these concepts provokes physical violence against trans people when the ‘deception’ is revealed: ‘A framework has been deployed whereby transphobic violence may be excused or justified on the grounds that deception had been involved’ (*ibid.*). Trans people are subject to these harms *not* as a permissible response to bad character traits or bad actions of theirs, but simply because of their gender identities, and no one ever deserves harm simply because of their gender. These harms are thus undeserved, and give us good reason to think that the concepts are defective.<sup>6</sup>

It is also worth noting that concepts that are themselves *not* defective can be misused, producing morally problematic effects similar to the above. Consider the misapplication of *thick* concepts—that is to say, concepts that have both descriptive and evaluative aspects, such as COURAGEOUS, COWARDLY, KIND, or CRUEL.<sup>7</sup> Applying a thick concept to someone involves some kind of evaluation of the person, and when that evaluation is unfair, the application of that concept might be wrongful. For example, Iris Murdoch describes the case of a mother-in-law, M, and her daughter-in-law, D. M was raised in an upper-class household with strong views about etiquette, so her initial impression of the lower-class and socially less restrained D is

---

6. To be clear, I do not presume to know *which* sex and gender conceptions are the ones we should be using. But the harms trans people suffer as a result of the widespread use of genitalia-based sex and gender concepts are grave and systematic enough that any conceptual engineering in this area must take trans experiences seriously.

7. Bernard Williams’s discussion of thick ethical concepts is the *locus classicus* on this topic (Williams 1985), though he acknowledges that the idea has earlier roots in the work of Ludwig Wittgenstein and those inspired by him, particularly Philippa Foot and Iris Murdoch. For a contemporary introduction to thick concepts, see Kirchin (2013).

unfavourable: she thinks of D as vulgar, undignified, noisy, and juvenile. Each of these concepts has a negative evaluative aspect; in thinking about D using concepts like VULGAR or JUVENILE, M thinks poorly of her daughter-in-law. However, M later subjects her characterisation of D to greater scrutiny, bearing in mind that her own elitist upbringing may have distorted her initial impressions. Through this reconsideration of the facts, ‘D is discovered to be not vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful, and so on’ (Murdoch 2013, 17). Murdoch argues that in coming to characterise D in different terms—that is to say, by using different *concepts*—M corrects a moral error: she manages ‘not just to see D accurately but to see her justly or lovingly’ (22).<sup>8</sup> In other words, it was wrong of M to think of D using those concepts, because they do not fairly apply to her. (If this seems difficult to accept, imagine how D, or her husband, might react if they were somehow to learn, before M’s reflection, what M really thinks of D. Naturally, they would feel insulted, and this implies that they would believe that M has done something wrong.) This need not suggest that these concepts are defective—the locus of the moral wrong lies in M’s *misuse* of those concepts in her thinking, not with the concepts themselves. Since the moral problems that can arise from concept use are broader than just those that are constituted by or that follow from the use of morally defective concepts, an account of conceptual responsibility should also capture cases like that of M and D.

---

8. In order to make it clear that the moral improvement here is entirely at the level of M’s concepts, Murdoch adds that M’s initial impression of D never showed in her behaviour: ‘the mother, who is a very “correct” person, behaves beautifully to the girl throughout, not allowing her real opinion to appear in any way...whatever is in question as *happening* happens entirely in M’s mind’ (Murdoch 2013, 17).

These examples suggest that the claim that there are morally defective concepts—and, in addition, that the misuse of benign concepts can be morally problematic—is true. But to answer *No Wrong Concepts* and establish a place for conceptual responsibility, we have to show how we could really be said to be doing something wrong when we use such concepts, if concept use is beyond our control. To do so, in the following three sections, I uncover and criticise the theory of moral responsibility that underpins *No Wrong Concepts*, and consider alternative accounts.

### 3. Voluntarism

For an agent to be responsible for something, that thing has to be connected to the agent in the appropriate kind of way, and the agent must not be excused from responsibility in some way. The classical view of moral responsibility—which I will call *voluntarism*—holds that the relevant connection is *voluntary action*.<sup>9</sup> For an action to be voluntary, it must meet two conditions. First, the agent must be in *control* of that action. Second, the agent must be *aware* of what she is doing in taking that action, and in particular, she must be aware of the moral worth of what she is bringing about. To take a fairly simple example, consider the following case:

*Teen Bully.* Lina is an 18-year-old delinquent who is mean to just about everyone. She is particularly harsh towards Gary, whom she regularly insults and beats up. Lina knows that it is wrong for her to be so mean to Gary, especially since Gary considers himself her friend. Yet, whenever Gary does something foolish, she ignores her conscience and bullies him anyway—after all, she finds it fun to be mean.

Because Lina is in control of her behaviour and aware that she is acting wrongly, she meets voluntarism's conditions for being responsible for bullying Gary.

When one or both of voluntarism's conditions are not fulfilled, the agent has an excuse that absolves her of responsibility. Potential excuses based on the *control* condition include

---

9. Voluntarism has too many adherents to acknowledge, going back at least to Aristotle (1999, especially 1110a–1111b4).

things like coercion, mind control, extreme duress, or (depending on who you ask) the truth of causal determinism. For example, consider the following variant on *Teen Bully*:

*Bully Puppet.* Lina bullies Gary whenever he does something foolish, and she knows that this is wrong. But, she is not in control of herself when she does so. An evil sorcerer has placed her under a curse, such that whenever Gary acts foolishly, she will be compelled to insult Gary and beat him up. Because of the curse's magic, Lina is completely powerless to stop herself from hurting her friend.

Because Lina is not in control of her behaviour in this modified case, she is excused from being responsible for bullying Gary.

Potential excuses based on the *awareness* condition include being unaware of or unable to comprehend what one is doing (e.g. a psychotic episode), or ignorance of morally relevant non-moral facts (e.g. not knowing that the tea one has served to a friend has been poisoned by a maleficent third party). Ignorance of the demands of morality (e.g. an ancient slaveholder who does not question the morality of the institution of slavery) may also provide an excuse, but this is contentious.<sup>10</sup> To take an example of non-moral ignorance, consider the following variant on *Teen Bully*:

*Ignorant Bully.* Lina has mended her ways and no longer consciously bullies anyone; indeed, she now intervenes to stop bullying whenever she sees it. One day, she happens across Gary and an acquaintance of theirs, Raz, who are fighting in an alley. Because Gary has the upper hand, Lina assumes that he was the instigator. She breaks up the fight by pulling Gary off of Raz, beating him back, and accusing him of having become a bully himself. Defeated, Gary apologises. But in fact, Raz was the instigator of the fight: he jumped Gary in the dark alley, and Gary was just defending himself.

Lina's intervention in the fight was unfair to Gary, and in fact turned into another instance of her bullying him, since she browbeat him into apologising for something he did not do. However,

---

10. See Rosen (2003, 2004) for an argument that moral ignorance can excuse wrongdoing. See Alvarez and Littlejohn (2017), Harman (2011), and Mason (2015), for arguments against the thesis that moral ignorance provides an excuse.

because she was ignorant of a morally relevant fact—that Gary was defending himself and not bullying Raz, as he appeared to be—she fails to meet the awareness condition, and is thereby excused. However, if Lina were somehow responsible for her ignorance—if her ignorance were the result of a wrongful act she voluntarily committed, say, refusing to listen to Gary’s attempts to explain himself as she accosts him—then her ignorance would no longer provide an excuse.<sup>11</sup>

Let’s turn now to the use of concepts. According to voluntarism, in order to be responsible for violating a moral standard by using a particular concept, the subject must both be in control of the use of that concept in her thinking, and aware it is morally problematic to use that concept in this context. These conditions severely constrain the range of cases in which the subject is responsible for using concepts in this way. We do sometimes actively choose to think using one concept rather than another—to think of something as *this* rather than *that*. In Murdoch’s story, for example, M might consciously try thinking of D as youthful rather than as juvenile, to see if this might be a more accurate, fair, and loving representation. But this reflective mode of thinking is not the typical way our concepts are deployed. Usually, when we acquire propositional attitudes, the concepts used to construct them are deployed through cognitive processes over which we have no direct control.

Moreover, just as it is uncommon to deliberately do something one believes to be morally wrong, it is rare to select a concept that one believes to be morally problematic.<sup>12</sup> Perhaps those

---

11. For more on the voluntarist account of culpable ignorance, see H.M. Smith (1983), Zimmerman (1997).

12. Except, perhaps, when one is being ironic or jocular, or when one is trying to interpret someone else’s thoughts or utterances in which a defective concept occurs, or when one is otherwise holding the offending representation at a distance. With regard to words, such contexts where one merely *mentions* an offending term are often taken to defuse the moral wrongs that typically follow from *using* the term, and we might want to extend this notion to the level of the concepts we employ. But see Langton (2018, 160), who argues that the use/mention distinction

who take glee in being ‘politically incorrect’ deliberately choose to use concepts like SAVAGE, knowing but not caring that it is wrong to dehumanise indigenous people. Similarly, some who deliberately apply genitalia-based sex and gender concepts to trans people may be aware that the entailed misgendering is undeservedly harmful. And, we might imagine a variation on Murdoch’s example where the mother-in-law deliberately continues to think unfairly of her daughter-in-law purely out of spite. But these kinds of cases will be uncommon. It seems far more likely that the thinkers in these cases would take themselves to be *in the right* by using those concepts in these ways. They would therefore fail to meet the awareness condition, and, thus, would not be responsible for using these concepts. Of course, they might well be responsible *for their ignorance* that using these concepts is wrong. For example, the racist’s ignorance regarding the wrongfulness of the concept SAVAGE may be due to a voluntary refusal to take indigenous perspectives seriously. But it still seems likely that such a refusal would be in *further* ignorance that it would be wrong to take the ignorance-inducing action. The agent would thus fail to meet the awareness condition at a higher level, and once again be excused from responsibility.<sup>13</sup>

According to voluntarism, the range of cases where we are responsible for using morally defective concepts is admittedly rather slim, but it is not empty, as *No Wrong Concepts* supposes. This might suffice to show that these concepts are indeed defective, because to use them

---

does not always work this way—in particular, the ‘injurious illocutionary potential of a slur’ can ‘leak’ out of the scare quotes one places around it.

13. This process of having to trace ignorant wrongdoing back through potentially multiple levels of ignorant wrongdoing to a clear-eyed wrongful act is the basis of Rosen’s (2003) argument that a voluntarist account of culpable ignorance leads to a kind of skepticism about moral responsibility. Though I will not take this up here, Rosen’s conclusion seems to me another good reason to reject voluntarism in favour of one of the alternative theories I present below.

voluntarily would be wrong. But this position would still be a disappointing result. It would mean that, in the vast majority of cases, we would not be permitted to *hold* people responsible for using morally defective concepts, because, according to voluntarism, we rarely ever *are* responsible for using a particular concept.

#### 4. Rationalism

The concern just raised touches on a more general problem for voluntarist accounts of moral responsibility, namely, that their central focus on *action* fails to account for the fact that we often hold people responsible for their *mental activity*, such as their propositional attitudes, what they remember or forget, what they notice or neglect, what occurs to them, and their emotional reactions. Angela Smith suggests the following example:

[*Birthday*.] I forgot a close friend's birthday last year. A few days after the fact, I realized that this important date had come and gone without my so much as sending a card or giving her a call. I was mortified. What kind of a friend could forget such a thing? Within minutes I was on the phone to her, acknowledging my fault and offering my apologies. (A.M. Smith 2005, 236)

The nature of Smith's moral fault in this case cannot be captured by voluntarism. She did not voluntarily choose to forget this important date, and probably couldn't have even if she wanted to. And yet, Smith continues, 'there was no doubt in either of our minds that I was, indeed, responsible for [forgetting]' (*ibid.*). Hence, Smith sought her friend's forgiveness for the fault.<sup>14</sup>

Noting that we in fact hold people accountable for the wrongs captured by cases like *Birthday*, Smith develops an alternative view of responsibility, which I will call *rationalism*. On Smith's account, the agent is responsible for behaviour of hers that reflects her *evaluative*

---

14. Readers unconvinced that forgetting to mark a friend's birthday *is* a moral fault may substitute their own example where one forgets something important, and thereby does something morally wrong—say, forgetting one's anniversary with one's spouse, or forgetting one's own *child's* birthday.



*outlook*, because it is that outlook that supplies the agent's *reasons* for what she does. As she puts it: 'an agent is responsible for  $\Phi$  just in case  $\Phi$  bears a rational connection to the agent's evaluative judgements' (A.M. Smith 2012, 577). The relevant notion of an evaluative judgement is broad, encompassing what we find true, important, valuable, or good, and going beyond consciously held attitudes to implicit and even unrecognised commitments. They are not merely one-off assessments, but rather 'continuing and relatively stable dispositions to respond in particular ways to particular situations' (A.M. Smith 2005, 251, fn. 27), which, 'taken together, make up the basic evaluative framework through which we view the world' (251). To summarise, according to rationalism, *the agent is responsible for  $\phi$ -ing if and only if the agent's reasons for  $\phi$ -ing stem from the agent's evaluative outlook.*

Rationalism captures the same cases as voluntarism, because voluntary action involves recognising and choosing to react to the reasons one takes oneself to have for that action, and these consciously recognised reasons often stem from one's evaluative outlook. For example, recall *Teen Bully*. Despite the fact that Lina knows she is doing something wrong when she bullies Gary, she may have a reason for doing so that stems from an objectionable part of her evaluative outlook, such as a judgement that the fun she gets from exerting her power over Gary outweighs the harm she causes him, or a judgement that her friends are not particularly important to her.

Rationalism also captures the standard kinds of excuses, but instead of appealing to the agent's lack of control over or awareness of what she is doing, the explanation for why she is excused points to the absence of rational influence from the agent's evaluative judgements. In *Bully Puppet*, for example, when Lina is compelled to bully Gary by the sorcerer's curse, her

actions do not reflect any of her evaluative judgements; the curse ensures that she acts entirely mechanically, without any capacity to respond to her reasons in action.

Cases like *Ignorant Bully* are slightly trickier for rationalism to handle. For it seems that regardless of her ignorance that she was doing something bad, Lina's actions reflect the rational influence of her evaluative judgements: in this case, Lina thinks that bullying is wrong and should be stopped, and for this reason she intervenes in the fight, unwittingly bullying Gary in the process. But here is one way the rationalist can make sense of why ignorance sometimes provides an excuse. When the agent is ignorant that she is doing something bad, this ignorance is itself something for which she may be responsible. So far, this is the same point about culpable ignorance recognised by voluntarism. But, according to rationalism, if the agent is *not* responsible for her ignorance, then *ipso facto* she is not ignorant because of the influence of any of her evaluative judgements. Something *other* than the agent's evaluative judgements thus accounts for the reason upon which she is *now* acting, because that reason stems from her ignorance, and the reason for her ignorance does not stem from her evaluative judgements. Therefore, she would not be responsible for acting badly.

In *Ignorant Bully*, for example, notice that although Lina's reason for intervening in the fight stems most directly from her evaluative judgement that bullying is wrong, this would be no reason to intervene had she not believed (falsely) that Gary was bullying Raz. This shows that her reason for intervening in the fight also stems from that prior belief. Now suppose that Lina is not responsible for being ignorant that Gary was just defending himself. This supposition entails that her evaluative judgements had no rational connection to her ignorance. So, her reason for intervening in the fight did not ultimately stem from her evaluative judgements after all. Therefore, she is excused for unwittingly bullying Gary.

Let's now return to how rationalism handles responsibility for mental activities. Smith's forgetfulness in *Birthday* can be interpreted as reflecting a judgement on her part that either her friend or celebrating birthdays is not particularly important to her. Smith's self-directed question—'What kind of a friend could forget such a thing?'—can be reformulated to ask for her reasons for forgetting—'Why did I forget her birthday?'—the answer to which points to the offending judgement—'I must not think she is all that important to me!'<sup>15</sup> Realising that she is responsible for this wrong, Smith then holds herself accountable by repudiating that judgement, admonishing herself for her error, and calling her friend to apologise.

Rationalism looks promising as a model for conceptual responsibility. Because it is not tied to the control condition of voluntarism, the fact that our typical use of concepts is not within our control poses no obstacle to the judgement that we may be responsible for using defective concepts, or for misusing benign concepts. On this view, what is required to be responsible for using a concept is that one's reason for using that concept stems from one's evaluative outlook. To illustrate, recall the examples I introduced in §2.

A subject who thinks of indigenous people as 'savages' would be responsible for using the concept SAVAGE, on this view, so long as the reason why she uses that concept is provided by her evaluative outlook. Suppose she believes some racist stereotypes about indigenous people—

---

15. This step of asking for the wrongdoer's reasons highlights that Smith's view is part of a cluster of closely related theories of responsibility (cf. Oshana 1997, Scanlon 1998, Hieronymi 2014) that all share a commitment to the idea that, fundamentally, what makes an agent responsible for something is that that the agent is *answerable* for it. That is to say, the agent is responsible for something when she is open, at least in principle, to demands that she explain her reasons for that thing. Smith's rationalist account is thus intended to provide the conditions under which the agent is open to this kind of demand. It is worth noting, however, that the notion of answerability can be accounted for by non-rationalist theories of responsibility. Sher, for example, whose view I discuss at length in §5, also refers to the notion of being answerable as the 'root meaning' of the term 'responsible' (Sher 2005, 68).

she thinks they are simple, violent, and anomic. As such she holds strong negative evaluative judgements about them and does not hold that respecting their humanity is important. These odious aspects of her outlook on the world would supply ample (bad) reasons for her to use the concept SAVAGE in her thoughts about indigenous people. She would thus be responsible for using this morally defective concept, and for the wrong she does to indigenous people by thinking of them as less than fully human.

In Murdoch's vignette, M's initially unfair characterisation of D reflects her old-fashioned and elitist worldview, from the perspective of which D's behaviour is improper. This judgement that behaviours like D's are improper gives M reason to use concepts like VULGAR and JUVENILE in thinking about D. But, as M's later reflection reveals, these concepts do not fairly apply to her daughter-in-law. Having realised this, perhaps M may go on to question those areas of her worldview. In any case, because M's initial commitment to these upper-class ideals is the source of her reasons for misapplying these concepts to D, she is responsible for thinking wrongly of her daughter-in-law.

The case of sex and gender concepts is somewhat more complicated. Unlike SAVAGE, which is defective in virtue of its morally objectionable content, concepts like MALE and FEMALE, or MAN and WOMAN, defined in terms of a person's genitalia, do not wear their defects on their sleeve. Rather, these widely-used sex and gender concepts produce harm because their content entails that trans people do not have the gender with which they identify, which in turn produces undeserved harms because this representation implies that trans people are 'pretenders' or 'deceivers'. In cases where the thinker holds an evaluative judgement that it is important to maintain that sex or gender is defined in terms of one's genitalia, and this judgement supplies the thinker's reason for using these concepts despite (or perhaps because of) these implications, the

thinker would be responsible for using these concepts and for any undeserved harms to trans people thus produced. However, it seems plausible that because many cis people (that is to say, people who are not trans) are ignorant regarding trans issues, many will thus not know that these concepts have harmful implications. In these cases, they would be responsible for using these concepts only if their ignorance is culpable. We may be able to trace their ignorance to a problematic evaluative judgement, such as a prejudice against trans people, in which case they would still be responsible for the harms produced by their use of these concepts. But in at least some cases their ignorance will be innocent, because their ignorance is traceable only to features of their situation, such as the paucity of information about trans experiences available in readily accessible media, and not to any problematic evaluative judgments. In these cases, their ignorance provides an excuse.

Finally, because not all of our thoughts reflect our evaluative judgements, sometimes the subject may not be responsible for using even a concept as clearly defective as SAVAGE. For instance, Smith specifically notes that rationalism rules out responsibility for ‘random thoughts and mental images’ that do not reflect one’s evaluative outlook (A.M. Smith 2005, 260). So when one has the familiar experience of an unpleasant thought ‘flashing’ across one’s conscious mind, one would not be responsible for the use of defective concepts in such thoughts. For example, suppose the refrain from ‘Savages’, a song from Disney’s 1995 animated feature *Pocahontas*, just starts playing, unbidden, in your head. Even though those thoughts may be constructed with a defective concept, the fact that these are just random thoughts and unreflective of any deeper evaluative commitment absolves you of any responsibility.

## 5. Psychologism

We might worry, however, about rationalism's reliance on evaluative judgements. For if behaviour for which we are responsible must be rationally connected to such stable and enduring aspects of our evaluative outlook, we would be unable to conclude that we are responsible for behaviour that is one-off, incidental, or out of character, no matter how heinous.<sup>16</sup> Recall that in *Birthday*, in order to vindicate Smith's assumption that she was responsible for forgetting her friend's birthday, we had to interpolate a bad evaluative judgement on her part—that she doesn't care enough about her friend, or that she doesn't think noting birthdays is worthwhile. While these interpretations are certainly plausible, it is equally plausible that Smith held no pernicious evaluative judgements at all. All of us forget important things from time to time. Sometimes, we do things we regret as a result. When this happens, it does not necessarily betray some malevolence or lack of concern that is hidden by our ordinary patterns of acceptable behaviour. Sometimes, we just screw up—but that does not necessarily shield us from responsibility.

This criticism of rationalism is similar to a familiar critique of a Humean view of blameworthiness, according to which the agent is to blame for bad actions of hers that stem from something 'durable and constant' in the agent, namely, her 'characters and disposition' (Hume 1896, 411). Like rationalism, a character-based theory of responsibility delivers unacceptable results when good people do bad things. For example, consider the following case:

*Good Teacher's Bad Day.* Sam is an adjunct professor. He is generally a decent person and an empathetic, supportive, and encouraging teacher who brings out the best in his students. But one day, stressed out by another failed round of applications for a permanent job, Sam comes into the classroom in a foul mood. His frustration manifests in his actions in class, and at one point he openly derides a question raised by an enthusiastic student.

---

16. I thank Miranda Fricker for discussion of this objection to rationalism.

We might sympathise with Sam—it is easy to act out when one is under stress and feeling dejected—but his behaviour is not, by that token, excusable. Yet, neither rationalism nor the Humean character-based view would allow us to conclude that Sam is responsible for mocking his student, because his reasons for this action did not stem not from a pernicious evaluative judgement or bad character trait, but rather, from a transient emotional state.

In response to the inadequacy of the Humean view, George Sher develops an alternative account. Sher argues that although Hume was wrong to focus on the agent's vices, he was right to ground responsibility in some of the agent's *constitutive psychological traits*. Building from this, on Sher's view, the agent is blameworthy for an action when 'the agent has failed to respond to some compelling moral reason for not performing it, and...the agent's failure to respond to that reason can be traced to the interaction of certain features of his situation with some significant subset of the desires, beliefs, and dispositions that together make him the person he is' (Sher 2005, 57). Because the desires, beliefs, and dispositions that cause the agent to fail to respond to the moral reasons he has are the same psychological traits which, taken together, are constitutive of who the agent is as a person, the badness of the agent's act reflects badly *on* the agent by way of those traits. This is so whether or not any particular trait is itself morally condemnable—no one element of the complex of the agent's constitutive psychological traits need be morally bad for some combination of them to cause the agent to fail to recognise her reasons against acting badly.

To illustrate, let's return to *Good Teacher's Bad Day*. Sam, by hypothesis, does not have any bad character traits, and, indeed, when it comes to his conduct in the classroom, he has traits we would call virtues. But on this particular day, his foul mood interacts with some combination of his constitutive psychological traits, causing him to fail to respond to his reasons not to mock

his student. The relevant traits might be quite fine-grained dispositions, such as a disposition to be curt and unfair when under extreme stress. Such a disposition is not obviously a morally bad trait, nor does it constitute a (mistaken) evaluative judgement that it would be appropriate to act in this way. Still, it is a part of the complex of psychological traits that make Sam who he is, and in this instance it causes him to fail to respond to his reasons not to mock his student. So, on Sher's view, Sam would be blameworthy for his bad conduct in the classroom.

Sher's view is intended to make sense of blameworthiness for actions, but it can readily be extended to cover responsibility more generally. The first modification to make is to note that our constitutive psychological traits do not just cause us to *fail* to respond to reasons *against* acting; they also cause us to *respond* to our reasons *for* acting. Second, recall Smith's point that many of our mental activities—our attitudes, what we notice or neglect, what we remember or forget, and so on—can be influenced by the reasons to which our evaluative judgements impel us to respond. Add to this Sher's observation that the *full range* of our constitutive psychological traits, and not just our evaluative judgements, has the potential to exert such rational influence. Putting these observations together, it is reasonable to think that the rational influence of our constitutive psychological traits also extends to our mental activities. In short, according to my proposed view, *the agent is responsible for  $\phi$ -ing if and only if the agent  $\phi$ 'd in responding or failing to respond to reasons she had for or against  $\phi$ -ing, where the agent's response (or lack thereof) to these reasons can be traced to the interaction of certain features of the agent's situation with some significant subset of the desires, beliefs, and dispositions that together make her the person she is.* This account, which I call *psychologism*, includes Sher's account of



responsibility for actions, but is generalised to capture responsibility for mental activities as well.<sup>17</sup>

To illustrate, recall our earlier examples. In *Teen Bully*, Lina would be responsible for bullying Gary because her failure to respond to her reasons not to do so can be traced to her general disposition to be mean to others. In *Bully Puppet*, she would be excused because the sorcerer's curse has disabled her ability to respond to reasons at all whenever Gary does something foolish, and therefore, her failure to respond to her reasons not to bully Gary cannot be traced to any of her constitutive psychological traits. Finally, in *Ignorant Bully*, Lina's failure to respond to her reasons not to bully Gary is due to her ignorance that Gary was defending himself from Raz's attack. If her ignorance, in turn, can be traced to some combination of her constitutive psychological traits, then she would be responsible for her ignorance, and therefore responsible for unwittingly bullying Gary. But if not, her ignorance would excuse her.<sup>18</sup>

With respect to Smith's forgetfulness in *Birthday*, some of the following psychological traits might explain her failure to respond to her reasons for remembering her friend's birthday: a disposition to not to remember dates, a desire not to make phone calls, a fine-grained disposition to turn inward and ignore others when (say) under stress, and so on. These psychological traits, supposing she has any of them, are parts of the complex that make her the person she is, and

---

17. I became aware, late in the writing of this paper, that Sher would likely resist this development of his account, as he has recently argued against the extension of morality into our 'private thoughts' (Sher forthcoming). I lack the space to engage fully with his arguments here, but I think my discussion of wrongful concepts in §2 might go some way to answering Sher's objections.

18. This description roughly follows Sher's own account of culpable ignorance, which see (Sher 2009, especially 85ff.).

cause her to fail to respond to her reasons for remembering her friend's birthday. So, according to psychologism, she would be responsible for forgetting her friend's birthday.

Before applying psychologism to concept use, it is worth considering an objection to the view. As Smith argues in a commentary on Sher's account: 'some sort of story has to be told about where the *activity* of the agent enters into this picture' (A.M. Smith 2008, 34, my emphasis), for it is such activity that explains what the view has to do with the agent's *agency*.<sup>19</sup> According to voluntarism, the agent's activity comes in via her voluntary *choices*. According to rationalism, the involvement of the agent's capacity of *judgement* implies some degree of activity, however non-voluntary, on the agent's part (A.M. Smith, 2005, 263). Where is the agent's activity on Sher's account, and on the view I have developed from it?

My response to this objection is that something very much like judgement, in Smith's sense, is involved whenever psychologism says that the agent is responsible for something. Recall that, according to Sher and to psychologism, the agent's constitutive psychological traits influence the agent's actions by way of the reasons to which they cause her to (fail to) respond. And *reasons-responsiveness*, like judgement, involves some activity on the part of the agent, however non-voluntary. The connection between the agent's constitutive psychological traits and what she does is thus not merely causal, in a purely mechanical kind of way, nor is the agent entirely passive with regard to this process. The agent's activity is involved because traits that are constitutive *of* her are what accounts for her reasons-responsiveness when she meets psychologism's conditions for being responsible for something.

---

19. I am grateful to Phyllis Pearson for pushing me to respond to this objection in a commentary she delivered on a related paper at the 2019 Congress of the Canadian Philosophical Association.

Finally, let's return to the examples of concept use. When the racist uses the concept SAVAGE in her thoughts about indigenous people, she fails to respond to reasons she has against using this concept—e.g. that it is dehumanising—because of one or more of her constitutive psychological traits—e.g. her factually and morally wrong beliefs about indigenous people. So, according to psychologism, she is responsible for using this concept.

When M misapplies concepts including VULGAR and JUVENILE to her daughter-in-law, she fails to respond to reasons she has against using these concepts in her thoughts about D—reasons that, upon reflection, she later recognises—because of one or more of her constitutive psychological traits—namely, her (at that point) uncritically endorsed upper-class ideals. So, according to psychologism, she is responsible for thinking of D using those concepts.

When the agent fails to respond to her reasons not to use genitalia-based sex and gender concepts—namely, because they can produce undeserved harms to trans people—and this failure is due to some of her constitutive psychological traits—such as a belief that genitalia-based sex and gender concepts should be retained despite these harms—the agent would, according to psychologism, be responsible for using those concepts. But, if the agent's failure to respond to those reasons was due to non-culpable ignorance of the harms these concepts can cause—i.e. ignorance that is not itself traceable to the agent's constitutive psychological traits—then the agent would be excused.

Lastly, when the agent has random thoughts that are constructed with defective concepts—such as when the refrain from the *Pocahontas* song starts playing in her mind—the agent would not be responsible for using these concepts, because these thoughts do not result from her failure to respond to a moral reason that was caused by her beliefs, desires or dispositions. Rather, the source of these thoughts is something sub-personal, mere mental

‘noise’. Such thoughts are not traceable to a failure to respond to reasons due to our constitutive psychological traits, but rather to processes in our unconscious minds that are not reasons-responsive at all.

## 6. Bad Conceptual Luck

While I argued in favour of psychologism on independent grounds, the upshot of the previous three sections is that we can find at least some cases, on each of the theories of responsibility that I presented, where the agent is responsible for using a concept. The claim that the agent does something wrong when she uses a defective concept is therefore vindicated, and *No Wrong Concepts* appears vanquished.

However, there remains a pressing objection to be addressed, which threatens to resurrect the conclusion of *No Wrong Concepts*. This objection concerns the role of *bad luck* in the formation of our conceptual repertoires. We acquire many, if not most, of the concepts we use via processes for which we are not responsible: our upbringing and early education, our acquisition of a natural language, our exposure to the surrounding culture. Even in cases where the agent otherwise meets the conditions of responsibility for using a concept, we might think that when the ultimate explanation for why she uses that concept is a matter of luck, she cannot fairly be held accountable for violating standards by using that concept.<sup>20</sup>

I have two responses to this objection. The first is that we in fact find many cases where holding someone accountable for violating a moral standard is warranted, even when the agent’s behaviour can be significantly attributed to bad luck. Consider the morally vicious. Suppose that

---

20. The problem of moral luck was introduced to contemporary philosophy in a pair of papers by Thomas Nagel (1979) and Bernard Williams (1981). I thank Miranda Fricker and Paul Faulkner for discussion of the luck objection.

Ivy Narcissa is the daughter of a multimillionaire, who raises her such that she develops multiple moral vices: arrogance, greed, and selfishness. Ivy's character development was thus largely a matter of luck, and not her own agency. Yet, she still seems open to criticism for her vices. Perhaps we would hesitate to blame her particularly harshly given the circumstances of her upbringing, but that should not make her immune to being called out in some way. Furthermore, the fact that Ivy's acquisition of these bad traits was largely a matter of luck does not shield her from being held accountable for bad *actions* she takes because her vices cause her to consistently fail to respond to certain moral reasons. Suppose Ivy attends a gala instead of visiting her sick friend in the hospital. Her friend would be right to criticise Ivy not just for her selfishness, but also for her selfish act, even though the ultimate explanation for her behaviour is the way she was raised.

In short, we hold people accountable for their vices and for the bad actions they take because of these traits, despite the their having these traits being largely a matter of luck. And so, the fact that our conceptual repertoires are also largely a matter of luck should present no obstacle to conceptual accountability. An agent who, for example, learned to use racist concepts such as SAVAGE because of the culture in which she was raised would still be accountable for violating moral standards by using this concept, when she meets the conditions for being responsible for her use of the concept.

Still, philosophers who resist the idea that bad luck could have any role in an account of responsibility (apart from serving as an excuse) may be unconvinced. Another approach—and this is my second answer to the objection—is to concede that we are not responsible when the ultimate explanation of our bad thoughts and actions is luck, while at the same time pointing out that we usually think we should *take responsibility* for these outcomes. Multiple philosophers

have defended this general approach. With regard to moral vices, Andrew Eshleman argues that we may not be responsible for acquiring these traits, but we should still take responsibility for them (Eshleman 2004). Susan Wolf has argued that a disposition to take responsibility for unlucky harmful outcomes of our actions is an ‘unnamed’ moral virtue (Wolf 2001). David Enoch goes further, suggesting that we have a moral duty to take responsibility for things connected to our identities but for which we are not responsible, such as the bad behaviour of our governments or close relatives (Enoch 2012). Finally, Iris Marion Young argues that even though we are usually not to blame for the unjust social structures in which our everyday and unobjectionable behaviour makes us complicit, we should all take responsibility these injustices insofar as we are able (Young 2011).

Well and good, but what, practically speaking, is meant by ‘taking responsibility’ for something? The kinds of actions involved depend on what one is taking responsibility for. They might include apologising, undertaking a project of self-improvement, revising one’s beliefs, assuming obligations to make amends for the outcome, or participating in political action. And, since we *should* take responsibility in these cases, it is plausible that others may press us to do so by reminding us of this requirement. In a sense, then, we may still hold people accountable in cases of bad luck, by urging them to take the actions that constitute taking responsibility.

Engaging in conceptual engineering can be viewed as a way of taking responsibility for our concepts. Critiquing our concepts for being morally defective, and developing revised or entirely new concepts that better serve our purposes, can and should go on even when we are not responsible for using them due to bad luck. As John McDowell argues, we have a ‘standing obligation’ to reflect critically upon our concepts, and so we must have ‘a standing willingness to refashion concepts and conceptions if that is what reflection recommends’ (McDowell 1996, 12–

13). Even if the fact that our conceptual repertoires are largely a matter of luck means that we are not responsible for using them, we should still hold ourselves and others accountable for our concepts by taking responsibility for them, and demanding that others do the same.

### **7. Conclusion: Conceptual Accountability**

We have now seen how we can answer *No Wrong Concepts* and establish the place of conceptual responsibility. On each of the theories of responsibility that I presented—voluntarism, rationalism, and psychologism—there is space to recognise that we may be responsible for using or misusing concepts. And, even if the fact that our conceptual repertoires are largely a matter of luck undermines this conclusion, we would still be required to take responsibility for our concepts. What remains to be said is what comes next: how should we hold one another to account for the concepts we use? By way of conclusion, I want to offer a sketch of conceptual accountability.

Moral philosophers often discuss moral accountability in terms of blaming, resenting, and punishing wrongdoers for their transgressions. These reactions serve not merely as catharsis for the blamer or deterrent for potential wrongdoers, but also to make the wrongdoer feel bad about what they have done, so that they recognise that they have acted wrongly. In other words, all blame implicitly or explicitly expresses some *criticism* of the wrongdoer. The most direct way to express this criticism would be through a speech act—e.g. ‘How could you think that was OK?’—but it can also be expressed non-verbally—e.g. storming out of the room, away from the wrongdoer.<sup>21</sup> When the agent is morally responsible for doing something wrong, she is open, in

---

21. This sketch of blame is heavily influenced by Miranda Fricker’s account of what she calls ‘communicative blame’, which on her view is the paradigm for all forms of blame (Fricker 2016).

principle, to being held accountable via these expressions of blame. Since, as we have seen, the agent is sometimes morally responsible for using or misusing concepts, it follows that, in these cases, the agent is open to blame and criticism for her concept use.

But recall that many of our concepts have defects that are not moral, but cognitive, theoretical, or semantic. While it is possible that using a concept with one of these kinds of defects would violate a moral standard—a bad theory can lead to harm in application—in many cases this will not hold—a person who conceives of the process of combustion using the concept PHLOGISTON may be seriously mistaken, but their use of this concept is probably harmless. In such cases, the agent who uses a cognitively, theoretically, or semantically defective concept would not be open to (moral) blame for doing so. Would we, then, not be permitted to hold people accountable for using such concepts?

There are difficult issues here regarding responsibility in domains other than the moral. While it is plausible that some version of the different theories of responsibility discussed earlier might be adapted to handle these cases, a full treatment must wait for another occasion.<sup>22</sup> For the moment it suffices to point out that we do, in fact, criticise mistakes in the cognitive, theoretical, or semantic domains. Fallacious reasoning, unjustified beliefs, flawed theories, and the use of nonsense terms all draw criticism that, like blame, aims to bring the person who made the mistake to recognise and correct the error. Sometimes, we even feel justified in resenting people

---

22. There have been several recent attempts to adapt theories of moral responsibility into the epistemic domain, which we might be able to employ with regard to responsibility for using cognitively, theoretically, or semantically defective concepts. Rettler (2018) argues for an account of doxastic responsibility based on a kind of indirect voluntary control over our beliefs. Hieronymi (2008) develops an account of responsibility for believing that is similar to rationalism. Some aspects of Sher's (2009) account of culpable ignorance might be extended to provide an account of responsibility for beliefs; Brown (forthcoming) modifies Sher's (2005) account of the nature of moral blame to provide an account of epistemic blame.



for their mistakes in these domains. So, a familiar way of holding people accountable is available to deploy when people use cognitively, theoretically, or semantically defective concepts.

By identifying and correcting defects in our concepts, conceptual engineering shows us where we need to hold each other and ourselves accountable in our thinking. But we cannot overlook conceptual engineers themselves—for conceptual engineering can be put to bad use, by philosophers or others. Just as concepts may be ameliorated to serve the aims of social justice or theoretical progress, much the same methods may be deployed to engineer concepts that are defective by design. We need only consider the concepts that extremist groups develop and modify to serve their ends, such as white supremacist conceptual inventions like RACE TRAITOR (a person who fraternises with people of a different race),<sup>23</sup> or terrorist distortions of traditional religious concepts like SHAHID (a kind of martyr in Islamic thought).<sup>24</sup> Conceptual accountability is not just concerned with the ethics of concept use. It must form part of the basis of a kind of applied ethics for conceptual engineering itself.

### References

- Alvarez, Maria, and Clayton Littlejohn. 2017. 'When ignorance is no excuse'. In *Responsibility: The Epistemic Condition*, edited by Philip Robichaud and Jan Willem Wieland, 64–81. Oxford, UK: Oxford University Press.
- Aristotle. 1999. *Nicomachean ethics*. Translated by Terence Irwin. 2nd ed. Indianapolis, IN: Hackett.
- Bettcher, Talia Mae. 2007. 'Evil deceivers and make-believers: on transphobic violence and the politics of illusion'. *Hypatia* 22 (3): 43–65. DOI: 10.1111/j.1527-2001.2007.tb01090.x.

---

23. Historically, 'race traitor' was also a term of abuse targeting white people who opposed racist institutions, such as Jim Crow laws in the US South or apartheid in South Africa. See Zack (1999) for discussion.

24. See Quiggin (2009) for discussion of this and other religious concepts that have been modified by Al-Qaeda to serve their agenda.

- Brown, Jessica. forthcoming. 'What is Epistemic Blame?' *Noûs* (published online Nov 2018): 1–19. DOI: 10.1111/nous.12270.
- Cappelen, Herman. 2018. *Fixing language: an essay on conceptual engineering*. Oxford, UK: Oxford University Press.
- Cappelen, Herman, and David Plunkett. forthcoming 2019. 'A guided tour of conceptual engineering and conceptual ethics'. In *Conceptual engineering and conceptual ethics*, edited by Herman Cappelen, David Plunkett, and Alexis Burgess, 1–27 in ms. Oxford, UK: Oxford University Press. Manuscript dated 30 July 2018, retrieved 9 July 2019 from URL: <https://philpapers.org/go.pl?id=CAPAGT&u=https%3A%2F%2Fphilpapers.org%2Farchive%2FCAPAGT.pdf>.
- Enoch, David. 2012. 'Being responsible, taking responsibility, and penumbral agency'. In *Luck, value, and commitment: themes from the ethics of Bernard Williams*, edited by Ulrike Heuer and Gerald Lang, 95–131. Oxford: Oxford University Press.
- Eshleman, Andrew. 2004. 'Responsibility for character'. *Philosophical Topics* 32 (1–2): 65–94.
- Fricke, Miranda. 2016. 'What's the point of blame? A paradigm-based explanation'. *Noûs* 50 (1): 165–83.
- Garcia, Ernesto V. 2002. 'A Kantian theory of evil'. *The Monist* 85 (2): 194–209.
- Harman, Elizabeth. 2011. 'Does moral ignorance exculpate?' *Ratio*, New Series, 24 (4, Special Issue: Developing Deontology): 443–68.
- Haslanger, Sally. 2012. 'Gender and race: (what) are they? (What) do we want them to be?' In *Resisting reality: social construction and social critique*, 221–247. Oxford: Oxford University Press.
- Hieronimi, Pamela. 2008. 'Responsibility for believing'. *Synthese* 161 (3): 357–373. DOI: 10.1007/s11229-006-9089-x.
- . 2014. 'Reflection and responsibility'. *Philosophy and Public Affairs* 42 (1): 3–41.
- Hume, David. 1896. *A treatise of human nature*. Edited by L.A. Selby-Bigge. Oxford, UK: Clarendon Press.
- Kirchin, Simon. 2013. 'Introduction: thick and thin concepts'. In *Thick concepts*, edited by Simon Kirchin, 1–19. Oxford, UK: Oxford University Press. DOI: 10.1093/acprof:oso/9780199672349.001.0001.
- Langton, Rae. 'Blocking as counter-speech'. In *New work on speech acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 144–64. Oxford: Oxford University Press, 2018.

- Mason, Elinor. 2015. 'Moral ignorance and blameworthiness'. *Philosophical Studies* 172: 3037–57.
- McDowell, John. 1996. *Mind and world*, 2nd ed. Cambridge, MA: Harvard University Press.
- Mills, Charles W. 2007. 'White ignorance'. In *Race and epistemologies of ignorance*, edited by Shannon Sullivan and Nancy Tuana, 13–38. Albany, NY: The State University of New York Press.
- Murdoch, Iris. 2013. 'The idea of perfection'. In *The sovereignty of good*, 1–44. London: Routledge.
- Nagel, Thomas. 1979. 'Moral luck'. In *Mortal questions*, 24–38. Cambridge, UK: Cambridge University Press. DOI: 10.1017/CBO9781107341050.005.
- Oshana, Marina A. L. 1997. 'Ascriptions of responsibility'. *American Philosophical Quarterly* 34 (1): 71–83.
- Quiggin, Tom. 2009. 'Understanding Al-Qaeda's ideology for counter-narrative work'. *Perspectives on Terrorism* 3 (2): 18–24.
- Rettler, Lindsay. 2018. 'In defense of doxastic blame'. *Synthese* 195 (5): 2205–2226. DOI: 10.1007/s11229-017-1332-0.
- Rosen, Gideon. 2003. 'Culpability and ignorance'. *Proceedings of the Aristotelian Society* 103: 61–84.
- . 2004. 'Skepticism about moral responsibility'. *Philosophical Perspectives* 18: 295–313.
- Scanlon, Thomas M. 1998. *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Sher, George. 2005. *In praise of blame*. Oxford, UK: Oxford University Press.
- . 2009. *Who knew? Responsibility without awareness*. Oxford, UK: Oxford University Press.
- . forthcoming. 'The wild west of the mind.' *Australasian Journal of Philosophy* (published online Sep 2018), 1–14. DOI: 10.1080/00048402.2018.1490326
- Smith, Angela M. 2005. 'Responsibility for attitudes: activity and passivity in mental life'. *Ethics* 115 (2): 236–271. DOI: 10.1086/426957.
- . 2008. 'Character, blameworthiness, and blame: comments on George Sher's *In praise of blame*'. *Philosophical Studies* 137: 31–39.
- . 2012. 'Attributability, answerability, and accountability: in defense of a unified Account'. *Ethics* 122 (3): 575–589. DOI: 10.1086/664752.

- Smith, Holly M. 1983. 'Culpable ignorance'. *The Philosophical Review* 92 (4): 543–571.
- Williams, Bernard. 1981. 'Moral luck'. In *Moral luck: philosophical papers 1973–1980*, 20–39. Cambridge, UK: Cambridge University Press.
- . 1985. *Ethics and the limits of philosophy*. London: Fontana Press.
- Wolf, Susan. 2001. 'The moral of moral luck'. *Philosophic Exchange* 31 (1): Article 1.
- Young, Iris Marion. 2011. *Responsibility for justice*. Oxford, UK: Oxford University Press.
- Zack, Naomi. 1999. 'White ideas'. In *Whiteness: feminist philosophical reflections*, edited by Chris J. Cuomo and Kim Q. Hall, 77–84. Lanham, MD: Rowman & Littlefield.
- Zimmerman, Michael J. 1997. 'Moral responsibility and ignorance'. *Ethics* 107 (3): 410–26.