

# 13

## Amelioration vs Perversion

*Teresa Marques*

C13

### 13.1 Introduction

C13.S1

C13.P1

Theories about the meaning of words for social categories and roles raise several interconnected questions. One is metaphysical: What is the nature of the social things we talk about? Another is conceptual: How do we think about the social world, and how does our thinking relate to the nature of social reality? And how should it relate? The latter question concerns so-called ameliorative projects. Such projects aim to address other additional questions, for instance, what is the point of having the concept in question? What concept (if any) would better serve our social or political goals? And who are *we*?<sup>1</sup>

C13.P2

In this paper, I argue that we should add one more question to this list, namely: How can we assess the legitimacy of ameliorative projects? I will try to answer this question by drawing from deeply problematic historical cases. Based on these cases, I'll introduce two notions of meaning perversion. This, I believe, will put us in a better position to understand the limits of the moral and political legitimacy of ameliorative projects. Permissible meaning revisions are those that are not meaning perversions. This does not tell us which are revisionary projects we should develop. But it gives us a way of circumscribing the meaning revisions we should not pursue.

C13.P3

Suppose we are interested in understanding crucial concepts of the language of justice and politics, e.g. in understanding the concepts expressed by words like 'democracy', 'freedom', 'fair elections', 'citizen', 'the people', etc. Sally Haslanger has persuasively made the case that we can mean different things by 'conceptual analysis', and that our reflection on the role of the language of politics can take different approaches. Haslanger (2006) distinguished between manifest, operative, and target concepts. A manifest concept is the concept one thinks guides one's categorizing, whereas the operative concept is the one that corresponds to actual categorization patterns. The target concept is the concept that, all things

<sup>1</sup> These questions are illustrated in debates about race, for instance in Appiah (2006), Glasgow (2003), Andreassen (2000, 2005), Kitcher (2007), Haslanger (2003, 2006), Machery et al. (2009), Diaz-Leon (2015), Ludwig (2018).

considered, we should employ given our interests, the facts, etc. (Haslanger, 2006, 99); it is the concept that, at the end of the ameliorative project, we should be using.

C13.P4 The question I address in this paper—*how do we assess the moral and political legitimacy of revisionary projects?*—is thus essential to know if a project is ameliorative. Haslanger considers a close question, and says ‘whether or not an analysis is an improvement on existing meanings depends on the purposes of the inquiry’ (Haslanger, 2012, n 1, p. 367). But I take it that it would be too naïve to believe that an analysis is an amelioration just in case the ‘engineers’ of the revision have good purposes. Some features of what Maynard and Benesch (2016) call Dangerous Speech, which I introduce below, show that people can be convinced of the goodness of a purpose while bringing out very bad outcomes. Hence, whether a revision is an improvement does not depend exclusively on how convinced theorists or activists are of the presumed goodness of their ends.

C13.P5 I take it, also, that there is some tension between the concept we should be using, all things considered and the concept we should be using, given our aims, interests, or the facts. The intersection of what we should do, all things considered, and what we should do, given our aims, leaves open the possibility that the answer to the legitimacy question may be epistemically unavailable, or that we can be wilfully ignorant of the right answer. It would thus be desirable to have some constraints or guidelines to assess a revisionary projects’ legitimacy.

C13.P6 Now, when theorists consider conceptual engineering or conceptual ethics, they are often concerned with the concepts a word should express, given the specific aims of the theory. Theoretical concerns are normally removed from the political domain. For instance, Tarski (1943) considered that ‘truth’ as used in natural languages is an incoherent notion, and that it gives rise to paradox. He argued that ‘truth’ should be defined in a way that would allow it to play its foundational role in a semantic theory, while insulating the theory from the contradictions and paradoxes that the natural language use gives rise to. For the purposes of the theory, a revision of the meaning of ‘true’ is desirable if the intended meaning fulfils the relevant theoretical aims.

C13.P7 Theoretical goals can also be pursued in social and political domains. Yet, there are differences between the theoretical domains that do not touch on socially relevant notions, and those that transpire into language use ‘in the wild’. Whereas the notion of truth in Tarski’s theory is not meant to replace the use of ‘truth’ in informal discourse, meaning revisions about social kinds or social relations often have both theoretic and social aims. For example, redefining what ‘marriage’ means in actual legislations has arguably changed the extension of ‘marriage’ in the real world.<sup>2</sup>

<sup>2</sup> I’m grateful to Pablo Rychter and Esa Díaz-León for discussion about this point.

C13.P8 This paper differs from other recent articles that try to pin down the normative constraints of conceptual engineering. For instance, Simion (2017) is optimistic about the prospects of conceptual ameliorations, but considers only epistemic constraints: a concept *should* be ameliorated, she claims, only insofar as this does not entail epistemic loss. Podosky (2018) argues contra Simion that epistemic loss is not an adequate criterion, and that conceptual revision should be allowed whenever it has the “capacity to causally influence the world”. Moreover, he suggests, normative questions should address feasibility constraints. He says that the claim that a concept should be engineered is right only if

C13.P9 . . . its desired causal influence on social reality is feasible; and the claim that a concept should be engineered is wrong if its desired causal influence on social reality is infeasible. Importantly, the notion of feasibility mentioned here is epistemic: It is about the feasibility of representational accuracy.

(Podosky 2018, 13)

C13.P10 In other words, conceptual engineering is feasible only if it is possible for a concept to come to accurately represent (social) reality. For instance, the revision of ‘marriage’ and the progressive change of legislation in liberal democracies around the world is feasible in Podosky’s sense, and this answers the question of whether ‘marriage’ should be ameliorated. This is an answer to one reading of the normative question.

C13.P11 This is not the reading of the ‘should’ in ‘when should a concept be ameliorated?’ that I have in mind.<sup>3</sup> I have in mind moral and political legitimacy. A conceptual engineering project could be feasible in Podosky’s sense, and still be morally and politically illegitimate. In other words, we should ask when are revisionist projects not ameliorative but perverse. I will argue that there is no obvious answer as to what ‘the value of a meaning revision’ can be beyond serving specific theoretical purposes, and that it is hard to balance the value of pursuing theoretical aims against the possible harms of using words ‘in the wild’, so to speak. The conviction that an analysis is pursued with good intentions does not guarantee the goodness of the outcome. And even when revisionary projects achieve their aims, the aims themselves may not be desirable. On the contrary, people should be cautious of deep feelings that their purposes are desirable or good. There may be cases where there is no clear answer to the question of what people should do. Given this, it would be desirable to establish guidelines for

<sup>3</sup> Lawford-Smith (2013) argues that the appeal to political feasibility is often used for ruling out political theories that can’t be implemented, but that it should rather be used as a tool for ranking alternative theories ‘along one of the dimensions relevant to making decisions about what to actually do’ (Lawford-Smith, 2013, 245). Furthermore, feasibility is independent of both desirability and risk. Hence, although understanding feasibility is an essential step in assessing political projects, it does not assess the desirability of the projects’ ends or risks.

assessing the moral legitimacy of revisionist projects. It would be useful to know, for instance, which projects *should not* be pursued. To this end, I'll consider two alternative definitions of meaning perversions.

C13.P12 Section 2 introduces some historical examples of meaning perversions, and points to cognitive and affective biases that are conducive to condoning harmful practices and behaviour, while protecting agents' moral self-righteous convictions. Section 3 explores the limits of our capacity to establish the moral and political legitimacy of revisionary projects, and takes race as a case study to illustrate the point. Section 4 characterizes meaning perversions. Ultimately, meaning perversions are revisionary projects that are politically or morally illegitimate. The section offers two different non-equivalent and non-extensional definitions of meaning perversion: one characterized in terms of harmful (perlocutionary) effects, and one characterized in constitutive terms. The possibility of meaning perversions in either sense establishes the limits of permissible meaning revisions.

C13.S2

## 13.2 Meaning Revisions in the Wild

C13.S3

### 13.2.1 Lessons from the Past and the Present

C13.P13 In *LTI, The Language of the Third Reich*, Victor Klemperer offers a chilling description of the corrupting power of language. Klemperer witnessed how, under the Third Reich, 'huge number of concepts and feelings' were corrupted and perverted. This was the case of words like 'heroic', 'heroism', or 'fanatic'. In his diaries, he registered how the use of these words shifted in a way that indicated that, for instance, one could not be a hero unless one were a fanatic. At the same time, fanaticism was no longer regarded as a negative trait. Klemperer reported that even after the Second World War ended,

C13.P14 . . . young people in all their innocence, and despite a sincere effort to eliminate the errors in their neglected education, cling to Nazi thought processes . . . as soon as this concept (heroism) was touched upon, everything became blurred in the fog of Nazism . . . and then replaced it with 'fanatical'.

(Klemperer, 1957/2000, p. 14)

C13.P15 He continues by drawing an illustrative analogy where Nazi language 'commanders for the party that which was previously common property and steps words, groups of words, and sentence structures, in poison.'

C13.P16 Masha Gessen has also written on the corrupting power of discourse. In a 2017 article, she warns us of the damage of Donald Trump's discourse to social and political reality. She compares the way Trump talks with the way autocrats in the former Soviet Union, and present-day rulers in Russia, talk:

- C13.P17 A Russian poet named Sergei Gandlevsky once said that in the late Soviet period he became obsessed with hardware-store nomenclature. He loved the word *secateurs*, for example. Garden shears, that is. *Secateurs* is a great word. It has a shape. It has weight. It has a function. It is not ambiguous. It is also not a hammer, a rake, or a plow. It is not even scissors. In a world where words were constantly used to mean their opposite, being able to call *secateurs* ‘*secateurs*’—and nothing else—was freedom.
- C13.P18 ‘Freedom,’ on the other hand, was, as you know, slavery. That’s Orwell’s *1984*. And it is also the USSR, a country that had “laws,” a “constitution,” and even “elections,” also known as the “free expression of citizen will.” The elections, which were mandatory, involved showing up at the so-called polling place, receiving a pre-filled ballot—each office had one name matched to it—and depositing it in the ballot box, out in the open. Again, this was called the “free expression of citizen will.” There was nothing free about it, it did not constitute expression, it had no relationship to citizenship or will because it granted the subject no agency. Calling this ritual either an “election” or the “free expression of citizen will” had a dual effect: it viscerated the words “election,” “free,” “expression,” “citizen,” and “will,” and it also left the thing itself undescribed. When something cannot be described, it does not become a fact of shared reality. Hundreds of millions of Soviet citizens had an experience of the thing that could not be described, but I would argue that they did not share that experience, because they had no language for doing so. At the same time, an experience that could be accurately described as, say, an “election,” or “free,” had been preemptively discredited because those words had been used to denote something entirely different. (Gessen, 2017)
- C13.P19 So, where Klemperer talks of ‘that which was previously common property’, Gessen talks of a lost ‘shared reality’. This is, I think, one of the harmful effects of meaning perversions, but here I will not elaborate further on what that amounts to, or how it opens the way for autocracy.
- C13.P20 I call these uses of ‘freedom’, ‘free expression of citizen will’, ‘election’, ‘heroic’, ‘fanatical’, *meaning perversions*. I consider two non-synonymous and (possibly) non-co-extensional senses of meaning perversion in Section 13.4: a causal consequentialist sense, and a constitutive sense. The next subsection illustrates the extent of the consequences of our cognitive failures and biases in some extreme cases.

C13.S4 13.2.3 The Road to Hell is Paved with Good Intentions

- C13.P21 The end aim of meaning revisions is sometimes self-interested. Some politicians’ use words like ‘law and order’ or ‘justice’ to manipulate the justice system in order to protect themselves or consolidate their power, not to promote the rule of law in

the service of justice and fairness. Other politicians' use words like 'democracy' and 'free elections' to perpetuate their hold on power, not to guarantee governments representative of citizen's rights and of the rule of law. Yet, meaning revisions often do have genuine social or political justice aims, or at least, aims that their proponents believe will bring about a better world. This section discusses some historical cases where extremely harmful consequences were seen as morally and politically permissible, or even required. People's inability to foresee those consequences were often due to common cognitive limitations. How does this relate to conceptual engineering projects? Conceptual 'engineers' may genuinely believe in the goodness of their purposes and methods. But there can be gaps between the best course of action for a certain end, theorists' assessment of to assess those means and end. And the fact that certain ends are *desired* does not make them *desirable* (all things considered), or desirable independently of the means set to achieve them.

C13.P22 In recent work, Maynard and Benesch (2016) discuss the conditions under which so-called Dangerous Speech (DS) can occur. They characterize DS as speech acts that are capable of encouraging approval of violence by an audience. This provides an extreme example of discourse that, from the perspective of the speaker and the target audience, is perceived as morally legitimate. As Maynard and Benesch characterize DS, it is a product of both its context and its content, which feed into and overlap with each other. The specific features of the contexts of DS are the (influential) speaker, the audience in its the socio-historical environment, and the availability of means of dissemination (Maynard & Benesch, 2016, p. 77). If a community relies mainly on one source of information, the message spread by that source has a greater impact. But an influential or authoritative speaker addressing a volatile audience through mass means of information is not producing DS if the content is not inflammatory or hateful. DS often occurs in social and historical contexts that increase the likelihood that the audience accepts that violence against certain people is morally permissible. It can be seen as rightful punishment for presumed past crimes, or as means to prevent presumed existential threats. Socio-historical contexts of DS may include longstanding grievances, resentment, the memory of historic injustices (real or imagined), a weak or dysfunctional justice system, competition for resources, or land disputes. Influential speakers may manipulate and exacerbate the resentment against members of another group for political gains.

C13.P23 Maynard and Benesch (2016) then offer a characterization of six possible features of the content of DS. These features need not all be present, and need not be present in the same way in all cases. The first is *dehumanization* through language,<sup>4</sup> which includes forms of discourse that can do direct harm by the

<sup>4</sup> See also Tirell (2012) or Jeshion (2016).

offense, denigration, or derogation of members of a target group as undeserving of the duties that are owed to them *qua* persons. The second feature of the content of DS is *guilt attribution*: members of a group are said to be guilty (as members of the group) of past crimes, e.g. rape or murder, stealing, responsible for current difficulties, destruction of the economy, occupation, or oppression, etc., and their guilt is offered as the moral justification for feelings of resentment and retributive action. The third feature of content is *threat construction*: an in-group accuses an out-group of posing an existential threat, which contributes to morally justify acting in self-defence. The fourth feature is the *destruction of alternatives* and the representation of a course of action as a historical necessity, or of alternative courses of action as impractical or inefficacious. For instance, Figs (2002) describes how many citizens of the Soviet Union thought that the violence of the Stalinist era was the only possible and necessary path to Communism, reporting someone that said: ‘I had my doubts about the Five Year Plan . . . but I justified it by the conviction that we were building something great . . . a new society that could not have been built by voluntary means’ (Figs, 2002, p. 111).

C13.P24 The fifth important feature is what Maynard and Benesch (2016) call *virtuetalk*, through which the audience is motivated by ‘deep and unreflected feelings that something feels “good” or “bad”, in particular feelings that induce positive moral self-appraisal’, a ‘satisfactory mental image of themselves . . . often shaped by notions of ideal group-identities, that produces considerable self-esteem’ (Maynard & Benesch, 2016, p. 84). It also contributes to the moral justification of actions against an out-group. The final feature of the content of dangerous speech is *future-bias*, i.e. the biased belief in a promise of future goods. Future-bias is presumed to outweigh the short-term difficulties the audience may have to endure, or the moral costs of the violence against others:

C13.P25 But the anticipated benefits can also be extravagant and utopian—promises that a positive transformation of society will be brought about through a temporary violent transition, or that national unity and prosperity for a long-mistreated people can be obtained. In light of the expectation that Soviet violence would protect the revolution and usher in Communist utopia, Lenin assured his followers that in the future ‘the cruelty of our lives, imposed by circumstance, will be understood and pardoned. Everything will be understood, everything’.  
(Maynard and Benesch, 2016, pp. 85–6)

C13.P26 DS thus arises when a perfect storm of cognitive biases and contextual conditions come together. The great motivational strength of DS—its nearly irresistible pull—comes from its reliance on moral reasoning and motivation. This motivational force arguably depends on the reactive attitudes involved. The notion of reactive attitudes was introduced by P. F. Strawson (1962/2008), and it includes attitudes like gratitude, resentment, and contempt. These are attitudes

that we have in reaction to another person's action towards us. They are essential both in interpersonal relations and in our moral lives.<sup>5</sup> Yet, our cognitive failures and biases—in ascribing guilt to others, in ruling out alternative courses of action, in seeing ourselves as virtuous, as *just knowing* that we are right, or towards some utopic future ideal, show that reactive attitudes and the actions they motivate can be easily misguided and morally unjustified.

C13.P27 Meaning perversions are not a minor issue. It should not surprise anyone to learn that meaning perversions played a central role in cases of DS that motivated historically inhumane acts of mass violence. Through meaning perversions, as Victor Klemperer put it, 'language does not simply write and think for me, it also increasingly dictates my feelings and governs my entire spiritual being the more unquestioningly and unconsciously I abandon myself to it'. Meaning perversions played a part in making it possible for Soviet citizens to tolerate and condone the actions that brought about the hunger of millions of people, for instance the Holodomor in Ukraine (a fact many still deny or minimize).<sup>6</sup>

C13.P28 How can we circumscribe the limits of morally and politically legitimate meaning revisions, in particular when politics or justice are at stake?

C13.S5

### 13.3 The Limits of Ameliorative Projects

C13.P29 This section advocates for defensive pessimism<sup>7</sup> in the assessment of the legitimacy of meaning revisions, i.e. in the assessment of the value and desirability of a meaning revision to a certain end. Revisionary projects should be approached not just with defensive pessimism, but also with honest humility. Indeed, theorists should not only be aware that any possible meaning revision may, and likely will, have harmful unintended consequences. Theorists should also be aware of the cognitive biases that often prevent them from seeing the harm their plans and actions bring about. As far as possible, such harm should be minimized. Below, the debate about the amelioration of the concept of race is introduced as a case study to illustrate the difficulty of circumscribing the politically relevant revisions that, ultimately, do no harm.

C13.P30 One way of deploying defensive pessimism, i.e. of taking proactive behaviour to counteract possible negative outcomes, is to adopt 'reflective equilibrium' as a method for finding coherence among diverse considered judgments, bringing them into relations of mutual support and explanation.<sup>8</sup> The process should be

<sup>5</sup> The notion has been deployed in recent moral and political philosophy, for instance by Manne (2017), Björnsson and Hess (2017), Goldman (2014), or Couto (2016).

<sup>6</sup> See Snyder (2017a, 2017b).

<sup>7</sup> The notion of defensive pessimism was introduced in Norem (2001). It is a cognitive strategy that helps people to take proactive behaviour to counteract possible negative outcomes.

<sup>8</sup> After Rawls (1971).

understood as allowing for the revision of convictions, given the facts and the foreseeable outcomes, a revision that honest humility can facilitate. But what are the morally or politically criteria that should guide our assessment of the facts and the foreseeable outcomes of concrete revisionary projects?

C13.P31 In a 2018 lecture, Haslanger argued against ideal theory and in favour of critical non-ideal theory to decide how to answer questions in social ontology—what possible and actual scenarios are relevant to consider. This decision, she argued, depends on what we want to know and for what purposes. Haslanger claimed that to address concrete justice issues—e.g. ‘How should we revise the educational system in Boston to be more fair?’—it is not helpful to start from idealized examples and ideal theory:

C13.P32 We can learn a lot about justice from considering concrete social circumstances. I would argue that such ‘bottom up’ (v. ‘top-down’) theorizing is the best way to learn about justice, for our background presuppositions are tested against the messy reality we are trying to address. (Haslanger, 2018, p. 2)

C13.P33 I don’t have a definite view here, but I am sceptical of doing away with ideal theory, and relying only on ‘bottom up’ theorizing. After all, without gaining distance from immediate practical concerns (interests and purposes) people may never move away from deep feelings that something feels right, while wilfully ignoring the harms they can cause by acting only on those feelings.<sup>9</sup> In the case of the rightness of a revision of meaning, the mere reliance on bottom up theorizing, and on feelings that something feels right, does not suffice to guarantee the moral or political legitimacy of a revisionary project.

C13.P34 Perhaps ‘the value of a meaning revision’ can refer to its positive overall consequences, all things considered. The distinction between an amelioration and a perversion would then depend on the balance of the good versus the harmful consequences of a revisionist project. But how can a theorist setting out to advance a new categorization know that good consequences will outweigh the harmful ones?

C13.P35 The opening section distinguished between meaning revisions that are pursued to meet specific theoretical aims and those that are intended to have real social consequences. This is a distinction that has been made in discussions about the usefulness of the concept of race. For instance, Kitcher (2007) defended a pragmatic biological view about race, considering that there are non-denumerably many ways to sort people into biological categories. These possible divisions, he claimed, depend on our cognitive capacities and our purposes. Other authors like Andreasen (2004) argued for a cladistic conception of human races (see also

<sup>9</sup> In any case, a reflective equilibrium approach can be pursued in a nonideal way. See, for instance, Stemplowska and Swift (2012).

Andreasen, 1998, 2000, 2005). One could perhaps ameliorate the concept of race to meet the purposes of a theory of human biology, and define races as ‘ancestor-descendant sequences of breeding populations that share a common origin’ (Andreasen, 2004, p. 425).

C13.P36 Yet, eliminativists like Appiah (Wilkins, Appiah, & Gutmann, 1998) argued that there are no human races, but not just because ordinary racial categories do not track any actual natural differences between people. The harms caused by the false belief in a unique biological base for racial categories is an additional reason to abandon all talk of human races. In European Portuguese, for instance, the term ‘raça’ is conatively loaded. Most people on the left avoid using the word. In fact, the end of the fascist dictatorship in Portugal in the 1970s led also to the independence of former Portuguese colonies in Africa (the revolution to end the dictatorship was carried out by the military who wanted to put an end to the war for the independence of those countries. The war had been going on for approximately thirteen years). In the new democratic constitution drafted by a mostly socialist and social-democratic parliament, all uses of ‘raça’ were removed. A similar process led to the removal of mentions of ‘Rasse’ from German legislation:

C13.P37 Recent debates about Weissein (‘whiteness’) in Germany provide further evidence that recognition of social realities of white supremacy does not presuppose an account of race in terms of these realities. For example, Amjahid’s (2017) book *Unter Weissen* (‘Among Whites’) dissects German practices of white supremacy but explicitly avoids realist appeals to Rasse by pointing out the ‘historical burden’ (2017, p. 49) of the concept. In this sense, the German concept of race may indeed be more adequately understood in analogy to other failed concepts such as *witch*. While alleged witches were forced in very real social positions, claims about the reality of races in Germany seem just as misleading as claims about the reality of witches. (Ludwig, 2018, p. 8)

C13.P38 In reply to similar concerns, Kitcher says that the usefulness of racial categories will depend on the theoretical purposes that are best served by having those categories.<sup>10</sup> But, as he puts it:

C13.P39 [E]ven if the concept of race plays a role in some lines of biological inquiry, the values of those lines of inquiry, and of pursuing them through retention of the concept of human race, would have to be sufficiently great to outweigh the potential damage caused by deploying this concept in the other contexts in which it plays so prominent a role, namely in our social discussions.

(Kitcher, 2007, p. 302)

<sup>10</sup> See also Díaz-León’s appeal to Kitcher’s criterion, Chapter 10 in this volume.

C13.P40 Kitcher's suggestion—that we can ameliorate the concept of race when the value of pursuing biological lines of inquiry is sufficiently great to outweigh the potential damage caused by deploying the concept in social contexts—seems a bit cavalier. The theorist must assume that the value of his inquiry justifies not only categorizing people into clades, i.e. ancestor-descendant sequences of breeding populations, which of course is useful for biology. The theorist is, on assumption, also using the word 'race' to refer to a clade, while knowing that millions of people have suffered because of categorization under different 'races', and that talking of biological races will continue to feed the social discrimination perpetrated on this basis. Redefining the meaning of race terms for the important goals of a theory does not do away, or minimize, the negative impact the use of 'race' has in society.<sup>11</sup>

C13.P41 What can we learn from this discussion? Can we say that the Germans and the Portuguese have ameliorated their race concepts by revising the terms used as empty, and hence removing them from legislation? That can't be right. Learning that a term has no referent is not the same as revising its meaning. 'Unicorn' does not refer to anything because there are no unicorns, not because people ameliorated the meaning of the word to give it a null-extension. Pointing out that there are species of mammals that may have given origin to the concept of unicorn—horses and narwhals—and that the study of these organisms is useful in biology, does not entail in any way that biology would benefit from using the term 'unicorn' to refer to either species. Equally, 'witch' does not refer to anything because there are no witches. Pointing out that there have been women who were considered inconvenient and who may have been at the origin of the concept *witch*—maybe those unmarried, maybe those with a mind of their own—and that our knowledge of their lives is useful historical knowledge, and perhaps a useful sociological warning sign of misogyny, does not entail in any way that social theories would benefit from using the term 'witch' to refer to actual women.

C13.P42 A meaning revision that gives an empty term a new extension cannot be ameliorative when giving the term *any* extension raises the probability of causing

<sup>11</sup> I don't address social constructionist non-eliminativism here, except briefly in a paragraph below. First, because *race* is not the main focus of this chapter. It is enough here to focus on what most people take races to be—biological kinds (see Machery & Faucher, Chapter 9 this volume, on the folk concept of race). Second, it is unclear that the most appropriate way of addressing social inequality is on the basis of a social constructionist non-eliminativist notion of race. Races are not obviously social constructions, unlike presidents, judges, professors, stock exchanges, etc. The claim that race is socially constructed is a theoretical view that social theorists can adopt to explain both the fact that it is not biologically real and that there are social relations (in several countries) that seem to be based on racial or ethnic group membership. But the point of treating race as a social construction is not just theoretical, it is practical—the theorist offers the ameliorated concept to replace the folk concept. But how do we know that deploying the theoretical notion of race in the real world would contribute to fairness? For all we know, it may be that reifying social identities is an obstacle to the creation of a fairer society. It may be that the social constructionist proposal faces the same challenges as the biological one: that redefining the meaning of race terms for the important goals of the theory does not erase the negative social impact of the use of 'race'.

very real harm. In the case of ‘race’, science can continue to research human reproduction and migration without using race-talk, as social sciences can without using ‘witch’. Race-talk has had, and continues to have, hugely harmful consequences. In the German or Portuguese contexts, the usefulness of pursuing lines of inquiry about human clades can be served with the technical term, ‘clades’, and not by the loaded terms ‘raça’ or ‘Rasse’. And, as Ludwig (2018) suggests, issues related to justice can be addressed instead by talking of racialized practices and conceptions.<sup>12</sup>

C13.P43 Now, even if it were true that there is only one way of classifying humans into different clades, it would still be the case that those divisions should not warrant racially based discrimination. For any biologically real characteristic, for instance biological sex, we should refrain from discriminating anyone on its basis. Discrimination differs from mere differential treatment.<sup>13</sup> This problem differs from the race problem discussed above. In the race case, there are good reasons to deny that there are races, because the underlying biological reality does not provide a coherent way of matching biological categories to pre-theoretical folk race conceptions, and race words are normatively loaded in highly pernicious ways. But we do track other biological categories fairly well, as is the case with sex categories, male and female, or conditions like albinism. Female people, or albino people, have suffered discrimination and harm for being what they are. That harm does not justify doing away with categorizing sexually male and sexually female organisms, any more than the harm that albino organisms suffer (they are prime targets for predators, for a start) does not justify doing away with the distinction between albinos and animals with (some) melatonin. Doing away with the categorization does not do away with the harm. When those organisms are human, recognizing them for what they are may be the required differential treatment that is necessary to address the injuries they are more prone to suffer. For instance, recognizing the differences between female and male symptoms of cardiovascular disease is necessary for properly diagnosing and treating females under forty who may die from heart attacks or strokes (Stamp, 2018).

C13.P44 In recent work, Haslanger (2019) offers a more sophisticated answer to the question of how to assess the adequacy of ameliorative projects. She draws from work by Knobe et al. (2013) and Knobe and Prasada (2011) on dual character concepts, i.e. concepts that have both a descriptive and a normative dimension, where dual character concepts

C13.P45 . . . are represented by (a) a set of concrete features and (b) a set of abstract values that the concrete features are seen as realizing. These two representations are

<sup>12</sup> This is also, in fact, the strategy for which Robin Andreasen argues (Chapter 11 in this volume), departing from her earlier view.

<sup>13</sup> See Lippert-Rasmussen (2013) for a discussion on ways to define discrimination.

intrinsically related, but they are nonetheless distinct, and they can sometimes yield opposing verdicts about whether a particular object counts as a category member or not. (Knobe & Prasada, 2011, p. 2965)

C13.P46 The dual character view offers a promising prospect for understanding the complexity of meaning change and will help to define one of the senses of meaning perversions in the next section.

C13.P47 This section showed that it is unclear what the value of a meaning revision is beyond the theoretical purposes it serves. The value of pursuing theoretical aims should be balanced against the possible harm caused by using the word in the wild. The conviction that an analysis is pursued to attain valuable theoretical or political aims does not guarantee the goodness of the outcome. On the contrary, theorists should be cautious of ‘deep feelings’ about the desirability of their projects and aims. Section 13.4 offers two definitions of meaning revisions that should not be pursued. It will take advantage of the notion of dual character concepts, and of the difference between harmful illocutionary effects and harmful perlocutionary effects.

C13.S6

### 13.4 The Legitimacy of Ameliorative Projects

C13.P48 This section introduces some theoretical resources that, together with the notion of dual character concepts, can help us understand language’s motivational power, and the illocutionary and perlocutionary effects that contribute to structure social relations. These resources are drawn not only from philosophy of language, but also from philosophy of action and of emotions. It makes use of these resources to distinguish between two senses of meaning perversion. Some of the notions introduced play a crucial role in the explanation of the expressive normative dimensions of meaning. Dual-character concepts reveal multidimensional aspects of meaning, which suggests that efforts to change one dimension of meaning may not be accompanied by changes in the other dimension.

C13.S7

#### 13.4.1 The Illocutionary Structure of Contexts

C13.P49 In recent joint work, García-Carpintero and I propose an expressive presuppositional account of derogatory language. Our view can be extended beyond derogatory language to other kinds of discourse that encode normative and evaluative expressive presuppositions. On our view, expressive presuppositions are not just propositions to be added to common ground as shared beliefs. Expressive language, e.g. language involving slurs, includes a cognitive and a conative dimension.

As we argue, expressives make requirements on a shared conative record, governed by *sui generis* norms specific to affective attitudes and their public manifestations (Marques & García-Carpintero, 2020).<sup>14</sup>

C13.P50 If emotions are *sui generis* normative states (Mulligan, 1998, D’Arms & Jacobson, 2000, Deonna & Teroni, 2015), and the speech acts that express them are defined by distinctive norms, then in order to incorporate the expression of emotions in a presuppositional view of pejoratives we should add further illocutionary structure to the context set. The intentional objects of the emotional states provide this additional structure. For instance, a pejorative or a slur presents its target as an adequate recipient of mistreatment, as worthy of contempt. The ‘formal object’ of the emotion is the normative condition that allegedly justifies the emotional attitude towards it. But this normative condition—the property of being contemptible—is not part of the represented content, and hence what speakers accept in accepting a use of a pejorative is not the belief that a target is worthy of contempt.

C13.P51 The nature of the expressive conative dimension of language use requires conversational contexts to have illocutionary structure,<sup>15</sup> including the different classes of contents to which speakers are committed in different modes: in the ways they are committed to their beliefs, to their intentions, to their affective attitudes, and to the questions guiding their inquiries. In felicitous contexts, these different commitments are mutually shared, and license presuppositions. As Stalnaker’s (1978) account of assertion emphasizes, an accepted assertion comes to be presupposed afterwards, allowing for the satisfaction of presuppositional requirements later on in the discourse. Similarly, accepted questions under discussion (QUDs), directives, and expressives can come to be taken for granted, constraining the legitimate moves that can later be made.

C13.P52 Our account fills in some of the details of the proposal made by Langton (2012):

C13.P53 I want to propose, in an exploratory spirit, the idea that the phenomenon of accommodation might extend beyond belief—beyond conversational score, and common ground, as originally conceived—to include accommodation of other attitudes, including desire and hatred. My remarks here will inevitably be programmatic. But to convey the general idea: just as a hearer’s belief can spring into being, after the speaker presupposes that belief, so too a hearer’s desire can spring into being, after the speaker presupposes the hearer’s desire; and so too a hearer’s hatred can spring into being, after the speaker presupposes that hatred. Stalnaker’s common ground can perhaps be extended to include not just common beliefs, and other belief-like attitudes, but common desires, and common

<sup>14</sup> See also Cepollaro and Stojanovic (2016) for a related hybrid presuppositional view.

<sup>15</sup> Charlow (2016) and Portner (2016) suggest that directives have a content to be added (when successful) to a collection of propositions that represent the mutually known active projects of the interlocutors, a ‘To Do List’ or ‘Plan Set’, and Roberts (2012) argues that contexts are further structured by the questions under discussion.

feelings, as well. Speakers invite hearers not only to join in a shared belief world, but also a shared desire world, and a shared hate world.

(Langton, 2012, p. 140)

C13.P54 The context update made with the literal use of an expressive is an update not only on the shared cognitive common ground—the set of propositions that come to be commonly accepted under a belief mode—but also on the shared ‘motivational set’<sup>16</sup>—the set of intentions, evaluative dispositions, or desires that are shared under intentional and sentimental modes.

C13.P55 A shared motivational set can be part of common ground, which would include those motivational attitudes that are in fact common, and taken to be such. But we can also represent a ‘motivational score’ as part of a conversational score, i.e. those attitudes whose expression is permissible in a context. The distinction is useful. People often accept to follow norms that they don’t agree with or that they disavow, and accept to follow widely shared and permissible evaluative attitudes that they do not actually have.

C13.P56 One may have doubts about the prospects of sharing emotions under a sentimental mode, even if the prospect of sharing plans does not raise the same doubts.<sup>17</sup> But authors like Salmela and Michiru (2016) have offered an account of collective emotions that links the intentional structure of joint actions to underlying cognitive and affective mechanisms. Collective emotions, they argue, can function as both motivating and justifying reasons for jointly intentional actions, in some cases even without prior joint intentions of the participants.<sup>18</sup>

C13.P57 Before embarking on meaning or conceptual revision projects, theorists should take into account the existence of dual-character concepts, as Haslanger proposes. But They should also consider the hybrid nature of evaluative and normative language, and seek a better understanding of the role of emotions in motivating and persuading through discourse. If words’ meanings can serve to make not only assertions, but also to make expressive speech acts, then the things (situations, events, or people) described can also be taken for granted as realizing certain norms or values through the emotional or affective attitudes expressed.

The next subsection shows how this hybrid nature of meaning can help us explain two kinds of meaning perversions.

<sup>16</sup> The suggestion adapts Bernard Williams’s (1979) notion of a ‘motivational set’, i.e. the set of dispositional attitudes, plans, intentions, emotions, etc., that identify the reasons agents have for acting, and are an integral part of their practical reasoning.

<sup>17</sup> On shared plans, see for instance Kutz (2000).

<sup>18</sup> Salmela and von Scheve (2017) deploy research on collective emotions to explain, for instance, right-wing radical populism, illustrating the usefulness of a notion of a common motivational set to explain, for instance, the functioning of Dangerous Speech.

C13.S8 13.4.2 Harmful Perlocutionary Effects or Constitutive  
Norm Erosions?

C13.P58 The main question this chapter addresses is how we can assess the legitimacy of ameliorative purposes and projects. The literature on conceptual engineering has not devoted sufficient attention to the possibility of meaning perversions. However, conceptual or meaning revisions can be either ameliorative or perverse. Ameliorations are improvements on existing meanings. Perversions, in contrast and by analogy, are corruptions of existing meanings.

C13.P59 This is still fairly imprecise, but it can be understood in two ways. First, perversions can be understood in causal terms. These could include many of the effects that Klemperer and Gessen mention: an impoverished experience, a destitute language, the loss of a shared reality, the loss of individual autonomy to a language that ‘thinks for us, and dictates our feelings’, as well as actual discrimination or oppression. The political dangers of these effects should not be neglected or minimized. These are often the effects autocrats intend, since they diminish a population’s capacity to resist the autocrat’s control over social and political reality:

C13.P60 *Causal meaning perversions* are attempts to hijack language (e.g. of justice, politics, social roles, or moral or epistemic virtues) in a way that makes people worse off. In other words, meaning perversions are revisionary projects that cause harm.

C13.P61 In a very literal sense, if a meaning revision produces harmful consequences it is not an amelioration. We may be concerned, however, that the causal notion of perversion cannot guide us in distinguishing between the merely feasible projects (in Podosky’s sense, introduced at the start of this chapter) and the morally legitimate projects.

C13.P62 We need a more perspicuous way of distinguishing ameliorations from perversions that goes beyond a focus on possible harmful effects. The reason is, fundamentally, epistemic. Shelly Kagan spells it out:

C13.P63 Perhaps the most common objection to consequentialism is this: it is impossible to know the future. This means that you will never be absolutely certain as to what all the consequences of your act will be. An act that looks like it will lead to the best results overall may turn out badly, since things often don’t turn out the way you think they will. Something extremely unlikely may happen, and an act that was overwhelmingly likely to lead to good results might—for reasons beyond your control—produce disaster. Or there may be long-term bad effects from your act, side-effects that were unforeseen and indeed unforeseeable. In fact, lacking a

crystal ball, how could you possibly tell what all the effects of your act will be? So how can we tell which act will lead to the best results overall—counting all the results? This seems to mean that consequentialism will be unusable as a moral guide to action. (Kagan, 1998, p. 64)

C13.P64 Kagan’s argument raises a problem for conceptual engineering projects generally.<sup>19</sup> It is a problem for efforts to discriminate ameliorations from perversions, and presses us to find a more useful way of identifying the engineering concepts we should be engaged in, morally and politically.

C13.P65 A second sense of meaning perversions, I suggest, is constitutive. The cases that Klemperer and Gessen focus on, together with the theoretical resources introduced in the previous subsection, and a modification of Jason Stanley’s definition of undermining propaganda (Stanley, 2015) provide the elements for a better definition.

C13.P66 Stanley defines undermining propaganda thus:

C13.P67 Undermining Propaganda: A contribution to public discourse that is presented as an embodiment of certain ideals yet is of a kind that tends to erode those very ideals (Stanley, 2015, pp. 52–3).

C13.P68 Now, I don’t use Stanley’s definition of undermining propaganda for two reasons. First, propaganda as he defines it will include uses of code words or dog whistles (Mendelberg, 2001, Khoo, 2017, Saul, 2018), or of racial fig leaves (Saul, 2017), and I think that meaning perversions differ in important ways from these other phenomena. Second, I think that meaning perversions can occur outside the remit of propaganda, for instance, in a relationship, where one of the parties can pervert the meaning of, say, ‘good friend’. Hence, not all propaganda is a meaning perversion, and not all meaning perversions are propaganda.

C13.P69 In my proposed definition, the idea of the undermining of norms and ideals is still essential. Typically, a correct use of word that expressively presupposes a norm or a value reinforces that value or norm. The norm or value is taken for granted as common ground. For example, thanking a child for being polite helps to reinforce polite interactions and rules of etiquette. Meaning perversions are uses of words that also presuppose shared norms or values, but pervert the norm enforcement process because they misapply terms or phrases to things that don’t realize the presupposed norm or value:

C13.P70 *Constitutive meaning perversions* A speaker *S* perverts the meaning of a word *w* just in case *S*’s use of *w* is presented as an enforcement or application of norms or

<sup>19</sup> This argument is made more precise in Lenman (2000).

values that *w* expressively presupposes, but erodes those very same norms or values by misapplying *w* to an unsuitable referent.

C13.P71 A use of a word is a perversion when it is false that the word applies to what the speaker intends to refer *and* the use of the word nonetheless has the illocutionary expressive effect it constitutively has—it expresses a conative state to the effect that what the speaker intends to refer to realizes a certain value.

C13.P72 Anyone who accepts the utterance making this expressive presupposition will come to take for granted that what is referred realizes that value. In other words, she will take for granted the permissibility of the relevant evaluative attitudes, and will possibly also share them. Those attitudes are often themselves motivating and justifying reasons for jointly intentional action. The use is perverse in that what is referred does not actually realize the presupposed value. And hence the motivated actions are not the appropriate actions to take towards the presumed referent.

C13.P73 What goes on with meaning perversions—like Putin’s ‘the dictatorship of the rule of law’, the Nazi ‘fanatic hero’, Medvedev’s ‘managed democracy’, Soviet elections as ‘free expressions of citizen will’—is that they take advantage of the hybrid contents these phrases encode. There are, on the one hand, legitimate referents of words like ‘democracy’, ‘free’, ‘elections’. The referents would be picked by the set of concrete features that correspond to one dimension of dual-character concepts. But there are, on the other hand, expressive normative or evaluative presuppositions that these phrases express. These would correspond to the set of abstract values those concrete features would realize.

C13.P74 Thus, a sincere literal use of ‘democracy’ made by a competent speaker denotes any form of political organization or government that displays some minimal features (allowing for more and less fitting cases—from full to flawed democracies). To accept that a state is democratic licenses certain presuppositions, not just about features of its form of government, but also about how it realizes certain desirable normative values. To take for granted that a form of government is democratic is to take for granted that it is desirable as good.

C13.P75 Besides the illocutionary effects, there are additional probable harmful perlocutionary effects of the use of a word *w*. The constitutive sense of meaning perversions and their perlocutionary effects are related. One sense in which we can pinpoint the nature of meaning perversions is that they have the harmful perlocutionary effects they have precisely because they contribute to undermine presupposed norms or values. But the two senses may not be co-extensional. Some revisions that have harmful consequences will not be constitutively perverse.

C13.P76 Take ‘free elections’: they are good things, they are essential for democracies, a recognition of citizens’ sovereignty through their representation in the institutions of their countries. The expression has a positive expressive connotation.

That positive aspect of the meaning of ‘free election’ can be taken for granted, at least for a good amount of time, to manipulate people. By describing the ritual practice of the so-called elections in the Soviet Union, the Soviet regime perverted the meaning of ‘free elections’. The ritual that was called a ‘free election’ did not display any of the concrete features that free elections must minimally exemplify. And hence, that ritual did not actually realize their positive normative value.

C13.P77 The normative presupposition expressed by ‘free election’, which is taken for granted as part of the motivational conversational set, pragmatically contradicts the actual application of the phrase to something that does not meet the minimal constraints for being a free election. By doing this, the Soviet regime eroded the positive value of ‘free election’, and normalized the new undemocratic practice. People lived under a pragmatic contradiction between the official normative ideal—democracies are valuable and participating in them is desirable—and the reality they were forced to inhabit. And people who were deprived of the means to appropriately describe the situation they lived were people deprived of the means to appropriately address it, and were more easily controlled by authoritarian regimes, as both Klemperer and Gessen testify. These are some of the harmful perlocutionary effects of meaning perversions.

C13.P78 Meaning perversions contrast with code words. Code words induce the acceptance into the motivational common ground of evaluative dispositions, plans, and norms, which conflict with pre-accepted dispositions/plans/norms that are part of a shared conversational score. But the expression of those values or norms is not encoded (not even as a presupposition) in the meaning of a code word.<sup>20</sup> This is exemplified in uses of code words lacking a negative racial connotation, for instance ‘we are doing a census of the people living in the *inner-city* to determine the investment in new and much-needed pre-schools and centres for primary medical attention’. Moreover, even those uses that do implicate a racial connotation denote their proper referent in the world, e.g. ‘inner-city’ refers to actual urban areas. That’s why plausible deniability is possible—one can always point out that what one said is true, and since the racist connotation is not part of encoded meaning, there is no contradiction in that denial.

C13.P79 In contrast, the normative and evaluative connotation of a meaning perversion is encoded in the word’s meaning, and thus it is automatically accommodated in the conversational record. Hence, speech that perverts meaning is easily accepted in a context since it aligns with what is already accepted as shared motivation. This explains Klemperer’s (1957/2000) description: ‘as soon as this concept (heroism) was touched upon, everything became blurred in the fog of Nazism’. However, meaning perversions are not used to denote their proper referents in the

<sup>20</sup> For discussion, see Khoo (2017).

world. And so, the evaluative connotation conflicts with the use of the word to talk about an improper referent that doesn't fit the value it is presented as realizing.

C13.P80 As a sign of these differences, the argumentative strategies of speakers who use code words and those who use meaning perversions differ. Code words allow plausible deniability: 'I wasn't saying anything about race! I was talking about criminality in certain urban areas'. Meaning perversions, in contrast, allow for rhetorical norm-enforcement questions: 'How can you be against our freedom?', 'How can you oppose democracy?', 'How can you believe what an enemy of the people says?' Interlocutors are, naturally, expected to reject that they are against freedom, or that they oppose democracy, or that they believe their enemies. They are also now pressed to accept that an improper referent has the concrete features that realize the values that they already take for granted.

C13.P81 It may be hard to know in advance if a given revision will have *harmful effects*, although we can try to fend them off for instance by combining, as I suggested, reflective equilibrium methods guided by defensive pessimism and honest humility. This means essentially that we should expect bad consequences to occur (and try to foresee them to the best of our ability) and be prepared to revise our expectations and plans if necessary. But it is not difficult to know in advance if a given revision is *harmful* if it is a meaning revision in the second, constitutive, sense. We can know it by understanding the mechanisms at play: intending to refer to an improper referent and accepting that it fits certain values or norms, which it very plainly cannot fit, is a perversion of those values.

C13.P82 When a revision is a perversion in the constitutive sense, it is especially difficult to resist: to give a reply to propagandistic calls for respecting the will of the people, for holding free elections, or for taking back control. Yet, in the mouth of many demagogues, these are meaning *perversions*: uses of 'the people' that exclude most of the people, of 'free elections' that are neither free nor an exercise in autonomous individual choice, of 'take back control' that give away control.

C13.P83 How do we resist meaning perversions, or spot them? Any direct criticism invites replies like 'how can you be against my freedom?', or 'how can you be against the people?' An interlocutor is left speechless, since in typical circumstances a normal reply would be obviously 'No! I'm not against the people, and I'm not against free elections'. Those rhetorical questions are effective to advance a meaning perversion because they seem to reinforce shared norms, while eroding them.

Conceptual 'engineers' who are interested in advancing a revisionist project can easily be unaware that they are putting forward a perversion instead of an amelioration. If, as Maynard and Benesch (2016) say, people are often moved by 'deep and unreflected feelings that something feels "good" or "bad"' that induce 'positive moral self-appraisal', conceptual 'engineers' may resist taking the extra step required to disentangle their conviction in the intended good results and the

appraisal that the misapplication of a word will erode the very same values or normative principles they believe they are promoting.<sup>21</sup>

C13.P84 This is a small step forward in delimitating the scope of the answer to the question of how we can assess the legitimacy of ameliorative projects. We can complement the discussion about why it is problematic to engineer race-talk, for instance. Recall that race-talk has had, and continues to have, hugely harmful consequences. Ordinary race concepts are presumed to track a set of concrete biological features that identify a natural class that turns out to be empty. The values that are supposed to be realized by those concrete features present some people as less deserving of consideration, social standing, or respect as persons by being presumed to exhibit certain biological features. We can try to ameliorate what ‘race’ refers—for instance, to intend it to refer to ancestor-descendant sequences of breeding populations that share a common origin. But the normative presuppositions that come with race-talk are not amenable to amelioration by *fiat*, particularly when there have been billions of people killed or enslaved on account of the negative values they were presumed to realize. The decision to do away with talk of *Rasse* or *raça* acknowledges this heavy burden. Moreover, there are alternative ways to address resilient racism, and to pursue useful lines of inquiry in biology, anthropology, or sociology.

C13.S9

### 13.5 Closing Remarks

C13.P85 I have argued for the importance of adding the question *How can we assess the legitimacy of ameliorative projects?* to the list of questions that conceptual engineering and conceptual ethics should address. I have also argued that our focus on the normative constraints of conceptual engineering should emphasize the moral and political legitimacy of the projects pursued. I was guided, in this regard, by troubling historical lessons. By characterizing meaning perversions, I hope to have offered a way of demarcating what a legitimate conceptual amelioration cannot be. The two senses perversion offered demarcate the set of morally permissible conceptual engineering projects. Revisionary projects that are permissible are the set revisions that are not perversions. That means that they are meaning revisions that (i) don’t have harmful consequences, and that (ii) do not misapply a word to something unfitting the values presupposed by its use.<sup>22</sup>

<sup>21</sup> Bicchieri (2005, 2017) work models how shared social norms that are followed in a society rely on people’s normative and empirical expectations with respect to what others will do and feel, and the conditions under which normative change occurs.

<sup>22</sup> Thanks to audiences at the Language and Power Workshop at the Eötvös University in Budapest in 2019, the Legal Theory Festival in Edinburgh in 2018, and in seminars and colloquia at the University of Barcelona, the University Pompeu Fabra, and the University of Valencia. Thanks in particular to Esa Díaz-León, Luis Duarte d’Almeida, Manuel García-Carpintero, Pablo Rychter, Jennifer Saul, Kathleen Stock, Kevin Toh, and Zsofia Zvolensky. Financial support was provided by

C13.S10

## References

- Andreasen, R. O. (1998). A new perspective on the race debate. *British Journal for the Philosophy of Science*, 49(2), 199–225.
- Andreasen, R. O. (2000). Race: Biological reality or social construct? *Philosophy of Science*, 67(3), 666.
- Andreasen, R. O. (2004). The cladistic race concept: A defense. *Biology and Philosophy*, 19(3), 425–42.
- Andreasen, R. O. (2005). The meaning of ‘race’. *Journal of Philosophy*, 102(2), 94–106.
- Andreasen, R. O. (2020). Conceptual fragmentation and the use of ‘race’ in scientific theorizing. In T. Marques & Å. Wikforss (Eds.), *Shifting concepts* (pp. 213–34). Oxford: Oxford University Press.
- Appiah, K. A. (2006). How to decide if races exist. *Proceedings of the Aristotelian Society*, 106(3), 363–80.
- Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.
- Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford: Oxford University Press.
- Björnsson, G. & Hess, K. (2017). Corporate crocodile tears? On the reactive attitudes of corporate agents. *Philosophy and Phenomenological Research*, 94(2), 273–98.
- Cepollaro, B. & Stojanovic, I. (2016). Hybrid evaluatives: In defense of a presuppositional account. *Grazer Philosophische Studien*, 93, 458–88.
- Charlow, N. (2016). Decision theory: Yes! truth conditions: No! In N. Charlow & M. Chrisman (Eds.), *Deontic modality*. Oxford: Oxford University Press.
- Couto, A. (2016). Reactive attitudes, forgiveness, and the second-person standpoint. *Theory and Moral Practice*, 19, 1309–23.
- D’Arms, J., & Jacobson, D. (2000). The moralistic fallacy: On the ‘appropriateness’ of emotions. *Philosophy and Phenomenological Research*, 61(1), 65–90.
- Deonna, J., & Teroni, F. (2015). Emotions as attitudes. *Dialectica*, 69, 293–311.
- Diaz-Leon, E. (2015). What is social construction? *European Journal of Philosophy*, 23(4), 1137–52.
- Figes, O. (2002). *The whisperers: Private life in Stalin’s Russia*. London: Penguin Books.
- Gessen, M. (2017). The autocrat’s language. *The New York Review of Books*. 13 May 2017.
- Glasgow, J. (2003). On the new biology of race. *Journal of Philosophy*, 100(9), 456–74.

the European Science Foundation, EUROCORES project CCCOM and FCT, Portuguese Government, project EuroUnders/0001/2010; by the DGI, Spanish Government, projects FFI2016-80588-R and FFI2015-73767-JIN; and by European Union’s Horizon 2020 programme under Grant Agreement no. 675415, project DIAPHORA.

- Goldman, D. (2014). Modification of the reactive attitudes. *Pacific Philosophical Quarterly*, 95(1), 1–22.
- Haslanger, S. (2003). Social construction: The ‘debunking’ project. In F. Schmitt (Ed.), *Socializing metaphysics* (pp. 301–25). Oxford: Rowman and Littlefield.
- Haslanger, S. (2006). What good are our intuitions? Philosophical analysis and social kinds. *Aristotelian Society Supplementary Volume*, 80(1), 89–118.
- Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford: Oxford University Press.
- Haslanger, S. (2018). Nonideal, critical, ameliorative, realist social ontology: Say what? Keynote lecture at *Social Ontology 2018: The 11th Biennial Collective Intentionality Conference*, 22–25 August 2018, Boston, MA.
- Haslanger, S. (2020). Going on, not in the same way. In D. Plunket, A. Burgess, & H. Cappelen (Eds.), *Conceptual ethics and conceptual engineering* (pp. 230–60). Oxford: Oxford University Press.
- Jeshion, R. (2016). Slur creation, bigotry formation: The power of expressivism. *Phenomenology and Mind*, 11, 130.
- Kagan, S. (1998). *Normative ethics*. Boulder, CO: Westview Press.
- Khoo, J. (2017). Code words in political discourse. *Philosophical Topics*, 45(2), 33–64.
- Kitcher, P. (2007). Does ‘race’ have a future? *Philosophy and Public Affairs*, 35(4), 293–317.
- Klemperer, V. (1957/2000). *The language of the Third Reich—LTI Lingua Tertii Imperii, A philologist’s notebook*. London: Continuum.
- Knobe, J., & Prasada, S. (2011). Dual character concepts. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2965–70). Boston, MA: Cognitive Science Society.
- Knobe, J., Prasada, S., & Newman, G. (2013). Dual character concepts and the normative dimension of conceptual representation. *Cognition*, 127(2), 242–57.
- Kutz, C. (2000). Acting together. *Philosophy and Phenomenological Research*, 61(1), 1–31.
- Langton, R. (2012). Beyond belief: Pragmatics in hate speech and pornography. In M. K. McGowan and I. Maitra (Eds.), *Speech and harm: Controversies over free speech* (pp. 72–93). Oxford: Oxford University Press.
- Lawford-Smith, H. (2013). Understanding political feasibility. *Journal of Political Philosophy*, 21(3), 243–59.
- Lenman, J. (2000). Consequentialism and cluelessness. *Philosophy and Public Affairs*, 29(4), 342–70.
- Lippert-Rasmussen, K. (2013). *Born free and equal? A philosophical inquiry into the nature of discrimination*. Oxford: Oxford University Press.
- Ludwig, D. (2018). How race travels: Relating local and global ontologies of race. *Philosophical Studies*, 176(10), 2729–50.

- Machery, E., Mallon, R., Nichols, S., & Stich, S. (2009). Against arguments from reference. *Philosophy and Phenomenological Research*, 79(2), 332–56.
- Manne, K. (2017). *Down girl: The logic of misogyny*. Oxford: Oxford University Press.
- Marques, T. and M. García-Carpintero (2020). Really expressive presuppositions and how to block them. *Grazer Philosophischen Studien*. 97 (2020), 138–58.
- Maynard, J. and S. Benesch (2016). Dangerous speech and dangerous ideology: An integrated model for monitoring and prevention. *Genocide Studies and Prevention: An International Journal*, 9(3), 70–95.
- Mendelberg, T. (2001). *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton, NJ: Princeton University Press.
- Mulligan, K. (1998). From appropriate emotions to values. *Monist*, 81(1), 161–88.
- Norem, J. (2001). Defensive pessimism, optimism, and pessimism. In E. Chang (Ed.), *Optimism & pessimism: Implication for theory, research, and practice* (pp. 77–100). Washington DC: American Psychological Association.
- Podosky, P. C. (2018). Ideology and normativity: Constraints on conceptual engineering. *Inquiry: An Interdisciplinary Journal of Philosophy* [online], 1–15.
- Portner, P. (2016). Imperatives. In M. Aloni and P. Dekke (Eds.), *Cambridge handbook of semantics* (pp. 593–628). Cambridge: Cambridge University Press.
- Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
- Salmela, M., & Michiru, N. (2016). Collective emotions and joint action. *Journal of Social Ontology*, 2(1), 33–57.
- Roberts, C. (2012) Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, 49(5), 1–69. Originally published in J. Yoon & A. Kathol (Eds.) (1998), *OSU Working Papers in Linguistics*, vol. 49.
- Salmela, M., & von Scheve, C. (2017). Emotional roots of right-wing political populism. *Social Science Information*, 56(4), 567–95.
- Saul, J. (2017). Racial figleaves, the shifting boundaries of the permissible, and the rise of Donald Trump. *Philosophical Topics*, 45(2), n.p.
- Saul, J. (2018). Dogwhistles, political manipulation and philosophy of language. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New work on speech acts* (pp. 360–83). Oxford: Oxford University Press.
- Simion, M. (2017). The *should* in conceptual engineering. *Inquiry: An Interdisciplinary Journal of Philosophy*, 61(8), 914–28.
- Snyder, T. (2017a). The deliberate starvation of millions in Ukraine: Review of Anne Applebaum’s *Great Famine—Stalin’s War on Ukraine*. *The Washington Post* 3 November 2017.
- Snyder, T. (2017b). *On tyranny: 20 lessons for the 20th century*. New York: Duggan Books.
- Stalnaker, R. (1978). Assertion. *Syntax and semantics* (New York Academic Press), 9, 315–32.

- Stamp, N. (2018). Women with heart diseases are dismissed and it's killing them. *The Guardian*. <https://www.theguardian.com/commentisfree/2018/jun/14/women-with-heart-diseases-are-dismissed-and-its-killing-them>.
- Stanley, J. (2015). *How propaganda works*. Princeton, NJ: Princeton University Press.
- Stemplowska, Z., & Swift, A. (2012). Ideal and non-ideal theory. In D. Estlund (Ed.), *The Oxford handbook of political philosophy* (pp. 373–89). Oxford: Oxford University Press.
- Strawson, P. (1962/2008). Freedom and resentment. *Proceedings of the British Academy*, 48, 1–25. Reprinted in *Freedom and resentment and other essays*, 2nd edn (pp. 1–28). New York: Routledge.
- Tarski, A. (1943). The semantic conception of truth: And the foundations of semantics. *Philosophy and Phenomenological Research*, 4(3), 341–76.
- Tirrell, L. (2012). Genocidal language games. In I. Maitra & M. K. McGowan (Eds.), *Speech and harm: Controversies over free speech* (pp. 174–221). Oxford: Oxford University Press.
- Wilkins, D. B., Appiah, K. A., & Gutmann, A. (1998). *Color conscious: The political morality of race*. Princeton, NJ: Princeton University Press.
- Williams, B. (1979). Internal and external reasons. In R. Harrison (Ed.), *Rational action* (pp. 101–13). Cambridge: Cambridge University Press.