Synthese (2021) 198:11587–11605
https://doi.org/10.1007/s11229-020-02815-9

Content internalism and conceptual engineering

Joey Pollock[1]

Received: 1 April 2020 / Accepted: 28 July 2020 / Published online: 5 August 2020

Abstract
Cappelen (Fixing language, Oxford University Press, Oxford, 2018) proposes a radically externalist framework (the 'Austerity Framework') for conceptual engineering. This approach embraces the following two theses. Firstly, the mechanisms that underlie conceptual engineering are inscrutable: they are too complex, unstable and non-systematic for us to grasp. Secondly, the process of conceptual engineering is largely beyond our control. One might think that these two theses are peculiar to the Austerity Framework, or to metasemantic externalism more generally. However, Cappelen argues that there is no reason to think that internalism avoids either commitment. Cappelen argues that to do so she must provide arguments for 3 claims: (a) there are inner states that are scrutable and within our control; (b) concepts supervene on these inner states; and (c) the determination relation from supervenience base to content is itself scrutable and within our control. In this paper, I argue that internalist conceptual role theories of content can meet Cappelen's challenge.

Keywords Conceptual engineering · Content internalism · Conceptual role · Semantic control

1 Introduction

Conceptual engineering is a methodology that consists in some combination of evaluating concepts or word meanings, identifying ways to improve them, applying these improvements, and/or implementing these changes in an individual or community (Cappelen 2018; Nado 2019; Eklund 2014). This might involve creating new or revised versions of representational devices or, in some cases, eliminating a word or concept altogether. There are many examples of conceptual engineering in both theory and practice. For example, we can understand the broadening of the concept MARRIAGE

✉ Joey Pollock
joeykpollock@googlemail.com

[1] ConceptLab, Department of Philosophy, Classics, History of Art and Ideas, University of Oslo, Postboks 1020, Blindern, 0315 Oslo, Norway

as an instance of ameliorative analysis:[1,2] the project is to improve the concept such that, whereas it was previously applicable only to heterosexual relationships, it should instead be broadened so that it can include partnerships between individuals of any sex/gender, such as same-sex partnerships. Similarly, one way to understand Clark and Chalmers' (1998) extended mind thesis is as a broadening of the concept BELIEF: whereas our traditional BELIEF concept applies only to states that are realized within the human body, the extension of the concept ought to be broadened such that, under the right conditions, it can include states supported by body-external devices such as smartphones and notebooks. Although a theorist could be interested solely in identifying which concepts we ought to employ, many conceptual engineers also believe that it is possible (and feasible) to successfully carry out plans to revise concepts in accordance with their ameliorative proposals. One commitment typical of engineers, then, is to treat the methodology of conceptual engineering as one that we can follow purposively; on this approach, concepts and word meanings are entities that agents can deliberately manipulate in stable and predictable ways. In what follows, I refer to this picture of conceptual engineering as the 'Autonomy View'.

In conceptual engineering projects that maintain the Autonomy View, there are identified benefits of implementing the new concept within a target community or individual, which benefits motivate the engineering of the concept. With respect to the amelioration of MARRIAGE, for example, these benefits are primarily moral and political: conceptual engineering contributes to efforts to achieve equality for the LGBT+ community.[3] For Clark and Chalmers (1998, p. 14), the benefits are primarily theoretical: the broader concept of belief is alleged to be more unified and more useful in explanation. The question of how to implement a new concept in an individual or community once it has been created introduces its own set of problems. These problems may be important to defenders of the Autonomy View: if one has created a new concept with the aim of securing some ethical or theoretical benefit, one will likely be interested in the question of how to implement this concept such that this benefit is secured (or, indeed, whether such implementation is possible at all). However, although these questions are relevant to understanding conceptual engineering projects undertaken by defenders of the Autonomy View, they are not my primary concern in this paper. My focus is on the question of whether concepts can be deliberately created or revised in predictable ways; I set the issue of successful implementation of revised concepts (in either an individual or a community) to one side.[4] I take the Autonomy View, then, to be concerned specifically with the issue of conceptual revision or creation, rather than the further issue of implementation. Having said that, I think that different views of content individuation and conceptual engineering may draw the line between

[1] The term 'ameliorative analysis' is from Haslanger (2000). I use the expression to mean the improvement of a concept by means of conceptual revision or replacement. However, it should be noted that, in some places, Haslanger describes amelioration as a process of uncovering which concepts we have been employing all along, rather than as one involving conceptual revision (Haslanger 2006).

[2] I use small caps to represent concepts and thought contents.

[3] For different analyses of the particular moral motivations for seeking marriage equality, see Calhoun (2002) and Richardson-Self (2015).

[4] For discussion of the issue of implementation, see Pollock (2019) and Koch (2018).

conceptual revision and implementation in different places; I return to this issue in Sect. 4.

Cappelen (2018) sets out a framework for understanding conceptual engineering—the 'Austerity Framework'—that entails a number of consequences that are incompatible with the Autonomy View. Conceptual engineering, according to Cappelen, is not the kind of thing that we can deliberately engage in to achieve predictable outcomes (although, he thinks, we will keep trying to do so). Cappelen's framework is based on a radical metasemantic externalism. One might initially think that it is only externalists that must reject the Autonomy View. However, pre-empting such a reaction, Cappelen poses a challenge to the internalist. He argues that there is no reason to think that internalists will have an easier time maintaining that deliberate conceptual engineering is possible—not until they provide arguments for claims concerning both our epistemic access to the facts that determine concepts and word meanings, and our ability to control such facts.

My aim in this paper is to meet Cappelen's challenge. I will not try to show that *all* forms of internalism can maintain the Autonomy View. Rather, I will focus on a solution to Cappelen's challenge for conceptual role theories in particular. I leave it open whether solutions are available to other internalist views (although I do note where my solutions have broader application). It may also be possible for some externalists to accept the Autonomy View. Koch (2018), for example, defends an externalist approach to conceptual engineering that maintains a significant degree of semantic control.[5] Whilst it is not my primary aim in this paper to argue against externalist approaches, I will offer reason for thinking that some kinds of internalist will have an easier time meeting Cappelen's challenge.

The paper proceeds as follows. In Sect. 2, I outline Cappelen's Austerity Framework; in Sect. 3, I set out his challenge to internalism. In Sect. 4, I introduce conceptual role theory and explain how the approach (or, some versions of it) can maintain the Autonomy View.

2 The austerity framework

Cappelen's (2018) Austerity Framework is a radically externalist approach to conceptual engineering. There are features of the framework that I will not cover here; I shall simply outline those that are important for my purposes. One thing that should first be noted is that the framework does not deal in concepts at all (2018, p. 61). Rather, for Cappelen, conceptual engineering aims to alter linguistic meaning (more specifically, the intensions and extensions of expressions). However, his challenge to the internalist is just as much a challenge for an internalist theory of concepts as it is a challenge to an internalist theory of linguistic meaning. In Sect. 4, I will focus on the

[5] See also Pinder (2019), who offers an account of conceptual engineering in terms of speaker-meaning that offers a degree of control over the process whilst remaining independent of any particular view of content individuation.

challenge as it applies to concepts and thought content; where relevant, I will explain how Cappelen's argument can be extended from word meanings to concepts.[6]

In the Austerity Framework, the supervenience base for meaning ('metasemantic base', in Cappelen's terminology) comprises a range of features familiar from the externalist theories of Kripke (1980), Putnam (1975), Burge (1979), and Williamson (1994). That is, meaning is determined, in part, by external factors that include facts about our physical environment as well as facts about how expressions are used, and have been used, in a community (2018, p. 62ff). These factors are also taken by many externalists to individuate concepts. Cappelen's Austerity Framework enthusiastically entails two controversial theses. The first he calls 'Inscrutable' (2018, p. 72):

> Inscrutable: The facts and mechanisms that underlie conceptual engineering are too complex, unstable, and non-systematic for us to grasp.

One reason that Cappelen maintains Inscrutable is that we don't have epistemic access to many of the facts in the metasemantic supervenience base—at least, not enough of it, and not for the majority of expressions. That is, we don't have much access to things like facts about past usage, introductory events, chains of communication, complex patterns of use across a language community, and so on. The problems don't end there, however. Cappelen points out that, even if we did have epistemic access to the facts in the supervenience base, it would not follow that we could know how changes to this base effect changes in meaning. To know this, we would also have to know something about the mechanisms of reference change. Cappelen argues that, even supposing that there is an algorithm that takes us from supervenience base to intension, we don't know what this algorithm is. Plausibly, we will never know this because it would be too complex for us to grasp. It may even be that no such algorithm *exists* (2018, p. 67). Thus, according to the Austerity Framework, we lack the knowledge that we would need to accurately predict how meanings will change based on our attempts to change them.

The second thesis is 'Lack of control' (2018, pp. 72–73):

> Lack of control: The process of conceptual engineering is largely beyond our control.

This thesis is, in part, a consequence of Inscrutable: we might be able to change some facts in the metasemantic supervenience base but, because of our epistemic limitations, we cannot do this in a way that would lead to predictable or controlled changes in meaning. In addition to this, Cappelen thinks that, even if we did have knowledge of the facts in the supervenience base, and of how changes to this base affect meaning, we would still lack any significant degree of control because this base includes many facts that we can't change (such as facts about the past) or that are extremely difficult to change (such as facts about patterns of use across a whole community). Both Inscrutable and Lack of Control carry over quite straightforwardly

[6] There may be differences in how the challenge confronts theories of word meanings in comparison with theories of concepts. My approach can be easily extended to word meanings if one thinks that word meanings just *are* concepts. Depending on one's view of the relationship between concepts and word meanings, however, Cappelen's challenge may be more problematic for word meanings. I do not deal with this issue here.

to the content of concepts to the extent that one thinks that concepts have the same externalist supervenience base as linguistic meaning.

The Austerity Framework is, thus, quite at odds with the Autonomy View. On the latter, conceptual engineering is a matter of purposively planning and applying changes to the metasemantic supervenience base that will result in predictable changes at the level of meaning and concepts. According to the Austerity Framework, in contrast, although we can make changes to the metasemantic supervenience base that will generate changes in meaning, an agent (or group) cannot make deliberate changes with predictable outcomes. Cappelen allows that an individual's plans (or beliefs, hopes, desires, etc.) regarding these changes may play a role in effecting semantic change, but they are just one small cog in a huge and chaotic machine whose inner workings are largely unknowable to us. According to Cappelen, the result is that, "you can engage in [conceptual engineering] without knowing that you are, you can think that you are doing it when you are not, and you do not know what changes to meaning you are making when you are making such changes." (2018, p. 73)

Initially, Inscrutable and Lack of Control may seem like the idiosyncratic commitments of a radically externalist metasemantics. More boldly, we might venture that this could contribute *support* for internalism: that is, if we think that providing a framework that makes conceptual engineering (of the sort envisaged by the Autonomy View) possible is a desideratum on a good theory of concepts, and that only an internalist could provide such an account (cf. Cappelen 2018, p. 82). However, Cappelen argues that this kind of reaction is misguided. In the next section, I set out his challenge to internalism.

3 Cappelen's challenge to internalism

Metasemantic internalism is the view that linguistic meaning and mental content, for an individual, are determined solely by factors internal to that individual. Internalism may appear to have an easier time avoiding commitment to Inscrutable and Lack of Control. This is because the factors that it appeals to—such as propositional attitudes or other mental sates—are factors that we might expect an individual to have a greater degree of power over. In this vein, Burgess and Plunkett write:

> The textbook externalist thinks that our social and natural environments serve as heavy anchors, so to speak, for the interpretation of our individual thought and talk. The internalist, by contrast, grants us a greater degree of conceptual autonomy. One salient upshot of this disagreement is that effecting conceptual change looks comparatively easy from an internalist perspective. We can revise, eliminate, or replace our concepts without worrying about what the experts are up to, or what happens to be coming out of our taps. From the externalist's point of view, however, conceptual revolution takes a village, or a long trip to Twin Earth. (2013, p. 1096)

Cappelen, however, argues that internalism is consistent with both Inscrutable and Lack of Control. There are three issues for the internalist. According to Cappelen, these each stem from the fact that internalism is a supervenience claim (2018, p. 82):

the view is that the meanings of expressions, for an individual, supervene solely on internal facts about that individual. The first issue is that, even if meaning supervenes solely on states of the individual, it might supervene on states that we do not have epistemic access to. Secondly, even if it supervenes *only* on introspectively accessible states, these might be states over which an individual has no control. Lastly, even if these states themselves are both introspectively accessible and under the individual's control, it may be that meaning supervenes on these states in unstable or unpredictable ways such that we do not know (and thus cannot control) how changes to these states will affect the meaning of an expression. At his most sceptical, Cappelen suggests that it is possible (for all we know) that meaning supervenes on internal facts such that it ends up being the *opposite* of what we had intended. He writes:

> even if there's supervenience on what we want or intend or decide, the supervenience relation doesn't have to make it the case that semantic values are what we intend for them to be, what we want them to be, or what we agree on them to be (for all we know, it could be a total mess or get us to the opposite of what we want, intend, or decide). (2018, p. 82)

Cappelen's argument is not that internalism entails Inscrutable or Lack of Control. Rather, what he argues is that there is no immediate route from internalism to the negation of these theses. The compatibility of internalism with the Autonomy View is something that the internalist must argue for. Cappelen, thus, challenges the internalist to provide arguments for the following claims (2018, p. 82): (a) there are inner states that are scrutable and within our control; (b) meaning supervenes on these inner states; and (c) the determination relation from supervenience base to meaning is scrutable and within our control. As above, this challenge carries over from linguistic meaning to concepts given that the problems concern the internalist metasemantic supervenience base. In what follows, I explain how one kind of internalist—the conceptual role theorist—can meet this challenge with respect to concepts. In the next section, I set out the view that I defend.

4 Internalist conceptual engineering

4.1 Conceptual role theory

Conceptual role theories claim that concepts are individuated by their relations to other concepts (and, perhaps, other kinds of entity) within a network.[7] For example, the concept CAT might be individuated, in part, by its relationships to the concepts, PET, ANIMAL, MAMMAL, and so on. We can represent concepts as nodes that occupy locations in the network and the relationships between concepts as links between nodes. On the particular kind of conceptual role theory that I defend, the relations that connect entities in the network come in a limited variety of primitive types. This approach follows work on network views of language in cognitive linguistics (e.g.,

[7] For conceptual role theories, see Rapaport (2002), Hudson (2007), and Field (1977). Closely related are inferential role theories—see Block (1986). For an externalist approach, see Harman (1987). Block's (1986) is a two-factor account that combines internal inferential role theory with a causal theory of reference.

Hudson 2007). One important kind of relation between concepts is that of category membership: concepts can be linked so as to form a hierarchy of subcategories and supercategories. For example, a subject's CAT concept may be related to her ANIMAL and PET concepts such that CAT is a member of the categories ANIMAL and PET (and not vice versa). Concepts can compose to form sentential contents. The relations between concepts (such as relations of category membership), determine which inferences sentential contents can participate in.[8] For example, if CAT stands in a relation of category membership to ANIMAL, then one will be disposed to infer from MARU IS A CAT to MARU IS AN ANIMAL. One will be disposed to draw this inference if and only if this relationship holds between the relevant concepts in the conceptual network.[9] Not all conceptual role theories are internalist; internalist theories treat the network as located in, or otherwise supervenient on, the internal states of the individual.

Conceptual role theories can be more or less holistic depending on whether it is the total network, or some part of it, that is responsible for individuating a particular concept. A radically holistic theory is one which claims that concepts are individuated, one-to-one, by their relations to *all* other concepts in the network. Radical holists must claim that a change to any part of the network determines a change in *all* concepts within that network; the result is that concepts (and, indeed, networks) cannot survive change. More moderate holisms claim that the relationship between total conceptual roles and concepts is many-to-one such that different (yet relevantly similar) conceptual networks might determine the same concepts (see e.g., Jackman 1999; Pagin 2006). Similarly, 'molecular' views claim that each concept is individuated by only a proper part of the total network. I think that both radical and moderate conceptual role theories can meet Cappelen's challenge. Where relevant, I will identify which aspects of this challenge might be more troubling for different versions of the view.

How does this approach accommodate conceptual engineering? Let's take the amelioration of BELIEF as an example.[10] In this case, amelioration begins with an individual identifying defects (or just room for improvement) in the original concept (e.g., 'bioprejudice', a lack of theoretical unity), and then proposing a successor concept, BELIEF* which lacks these defects (e.g., it is inclusive of states realised in body-external resources). In ameliorative analysis, this successor concept is to be expressed with the same word-form ('belief') to exploit the lexical effects associated with the term (Cappelen 2018). The specification of this successor concept generates a new node in a particular location in the conceptual network, with links to existing nodes that delimit its conceptual role. Thus, on this view, a conceptual 'revision'

[8] As I will argue in Sect. 4.2, this presentation of the view may need to be complicated to accommodate cases in which the agent understands which inferences are licensed by a concept and yet is not disposed to draw these inferences.

[9] This is a simplified presentation of the view. One issue that must be dealt with is the apparent existence of cases of non-monotonic inference. For example, supposing that the concept CAT is a subcategory of FURRY ANIMAL, how does the account deal with the classification of hairless cats, such as the sphynx? A response to this problem is offered in Hudson (2007, p. 25ff).

[10] A moderate conceptual role theorist might allow that different agents can possess tokens of the same shared concept—e.g., BELIEF. A radical holist, in contrast, will not think there is such a thing as *the* concept BELIEF. Rather, individuals each possess their own idiosyncratic concepts which stand in varying degrees of similarity to each other. I ignore this complication in what follows.

typically involves the creation of a new, type-distinct concept.[11] The engineering process is one that acts on concepts in an agent's internal conceptual network; however, importantly, one does not have to conceive of oneself as updating a conceptual network in order to engage in conceptual engineering (just as one does not have to conceive of oneself in this way when learning a new concept). The conceptual role theorist's claim is that the conceptual network constitutes the cognitive underpinnings of natural language understanding—a claim that, as Hudson (2007) notes, has significant psychological plausibility.[12] Any change in concepts, or in use of expressions, must be underwritten by changes to the conceptual network. The self-conscious mental or linguistic behaviour that an agent engages into effect such changes, however, may take a variety of forms. For example, agents may devise full or partial definitions, select exemplars or paradigms (e.g., Otto's notebook), make decisions regarding how to classify worldly objects, or specify patterns of inference. Changes in any of these behaviours necessitate corresponding changes to the conceptual network.[13] Similarly, implementing a concept in a community or individual requires (amongst other things) generating the same or similar changes to the conceptual networks of other agents, where network changes can be produced by convincing agents to alter these same sorts of linguistic or mental behaviours.

Beyond this initial sketch of the approach, there are further choice points regarding how to understand conceptual engineering as a conceptual role theorist. These will emerge in the course of my response to Cappelen's challenge. I turn next to this response.

4.2 Meeting Cappelen's challenge

4.2.1 There are inner states that are scrutable and within our control

The first claim that the internalist must maintain is that there are inner states that we have epistemic access to and that we can deliberately manipulate. Let's take the epistemic issue first. Initially, this may seem straightforward for the internalist: it is typically thought that internalism allows that content is 'transparent' to the agent and, thus, an internalist should have no problem claiming that we have epistemic access to content. In contrast, externalists are thought to be committed to rejecting transparency.

[11] One could perhaps think of the ancestor and successor concepts as two versions of the same concept. However, for the purposes of this paper I will assume that the successor concept is type distinct from its ancestor.

[12] For an overview of empirical evidence supporting the existence of a language network, see Hudson (2007, p. 36ff). It should be stressed that the psychological plausibility of the language network does not settle any metaphysical questions regarding the foundations of meaning and concepts. Whereas an internalist may take this network to be solely responsible for individuating concepts, an externalist may treat this network as simply realising our subjective understanding of, or epistemic relationship to, externally individuated concepts.

[13] One consequence of this is that, although a conceptual engineer may not endorse conceptual role theory, or even the Autonomy View, so long as she engages in these sorts of mental and linguistic behaviours, or causes them in others, she will be generating network changes and thus altering concepts. Haslanger (2000), for example, in setting out her proposals regarding gender and race concepts and convincing others with her arguments, will thereby be generating changes in both her own, and other agents', conceptual networks.

Whilst I think that internalists can indeed maintain transparency, I also think that the issue is not as straightforward as it may first appear. In what follows, I explain why this is the case and identify the kind of epistemic access claim that the conceptual role theorist should endorse.

Following Wikforss (2015), it is helpful to separate two kinds of transparency that are each invoked in the internalism/externalism debate. One notion, 'access transparency', concerns an individual's epistemic access to the contents of her thoughts. Content is access transparent to the agent if she can tell, through introspection, when two concepts or thoughts are the same and when they are different (e.g., Boghossian 1994). As Wikforss explains, this kind of transparency requires the ability to form (true) beliefs about meaning and content on the part of the agent. The second kind of transparency is 'functional transparency' (Recanati 2012). This kind of transparency is non-epistemic. It concerns the determination of content and its role in reason and action. Roughly, content is functionally transparent if it captures the agent's cognitive perspective. This does not require the ability to form metaconceptual beliefs. As Wikforss describes:

> functional transparency does not concern access to contents but is just the thesis that thought content is determined by cognitive role. Thus, if *S* reasons as if two thought tokens are about distinct objects […] then the thoughts have distinct content; conversely, if *S* reasons as if two thought components have the same content […], then they do. (2015, p. 147)

Externalists cannot maintain either notion of transparency, although they may be able to maintain related weaker theses, such as the thesis that an agent can have self-knowledge in Burge's (1988) sense. Depending on their view, an internalist might hold both transparency theses, or they may hold only the weaker functional thesis.

Which transparency thesis is relevant to Cappelen's challenge, and can the conceptual role theorist maintain it? Internalist conceptual role theories can easily maintain functional transparency—indeed, functional transparency is essentially just the thesis that internal conceptual role determines content (cf. Wikforss 2015, p. 159). That conceptual role theorists can maintain functional transparency is, however, not obviously enough to meet Cappelen's challenge. This is because the process of deliberately engineering concepts seems to require beliefs about content—for example, the belief that some concept, C, is defective, and should be revised or replaced. Conceptual engineering, thus, seems to require some kind of access transparency.

The apparent need for access transparency presents the conceptual role theorist with potential difficulties. Some versions of the access transparency thesis are controversial, to say the least. As Wikforss argues, if access transparency requires direct introspective access to concepts conceived of as symbols or vehicles in the language of thought, the claim looks implausible.[14] She writes that "the very idea that knowledge of content involves having introspective access to the content of mental expressions would seem deeply problematic." (2015, p. 153) Moreover, as she points out, prominent defenders of the language of thought hypothesis have denied that we have this kind of access

[14] Wikforss (2015) is not arguing against all forms of transparency. Rather, she is arguing that access transparency does not have the semantic significance that it is typically treated as having in the internalism/externalism debate.

(Fodor 1975). How, then, should a conceptual role theorist understand epistemic access for the purposes of conceptual engineering? I do not think that the conceptual role theorist need claim anything so strong as that we have direct introspective access to mental vehicles. What Cappelen's challenge requires is that we can form true beliefs about concepts such that we can reliably track the changes that we make to them. But there are, in principle, multiple ways that a subject could form such reliable metaconceptual beliefs other than through direct introspection of mental vehicles. To take an unrealistic, yet illustrative, example, if I have an oracle that regularly and reliably informs me of the contents of my thoughts, I will have the kind of epistemic access that Cappelen's challenge requires: I will be able to track how changes to the metasemantic supervenience base effect changes to my concepts. Now, obviously, conceptual engineers do not have oracles at their disposal. However, as Wikforss (2015, p. 163) also maintains, the possibility for the reliable formation of metaconceptual beliefs can be secured by appeal to functional transparency: because conceptual role determines content, our inferential practices will be a good guide to the metaconceptual truths. Functional transparency does not itself involve any metaconceptual capacities, but it does make our metaconceptual judgments reliable. For example, if an agent is disposed to infer from STANLEY IS AN OCTOPUS to STANLEY IS AN ANIMAL, this justifies her in forming the belief that the concept OCTOPUS is a subcategory of the concept ANIMAL because she would not be disposed to draw this inference unless these concepts stood in this relationship to each other. This is not to say that agents can never make any inferential or metaconceptual mistakes. But an agent can, perhaps not infallibly, but reliably, judge which relations concepts stand into one another.

Conceptual role theorists can, then, maintain that we have epistemic access to concepts in the conceptual network. Do we have control over this network, though? This issue is also not straightforward. What exactly does it mean to say that we can control our concepts? How much control is needed for conceptual engineering? Recent research on mental control suggests that agents do not exhibit very much control or autonomy over their own mental states. Metzinger (2013), for example, argues that we exhibit mental control during only about a third of our lives. This sounds potentially bad for the internalist; however, as I will argue, a closer look at Metzinger's notion of mental control will give us reason for optimism. In what follows, I will argue that conceptual role theorists can claim that we have a weak sort of control over our concepts. And, importantly, this degree of control is enough for engineering purposes.

Metzinger (2013) offers an account of what it is to have control (and lack control) over mental activities. Mental control (mental 'autonomy', in Metzinger's terminology) is the ability to control one's own mental functions in a way that is consciously goal-directed.[15] Mental control is exhibited when we perform various kinds of mental action. Metzinger lists as examples of autonomous mental actions, control of attention, episodic memory, planning, rational deliberation, decision making, and concept formation (2013, p. 2). He identifies two kinds of mental action and agency—attentional and cognitive. Attentional agency is the ability to control one's attentional focus, as when one chooses to attend to a particular object; cognitive agency is the ability

[15] Metzinger's account is independent of metaphysical disputes regarding free will. His aim is to demarcate a kind of mental activity that has distinctive functional and phenomenological features.

to control goal-directed deliberative thought. Concept formation plausibly falls into this latter category—it is intentional mental action directed at the goal of creating a new concept. Autonomous mental actions, unlike various unintentional mental activities, are distinctive in that they can be deliberately inhibited, suspended or terminated (2013, p. 3). Mental control is contrasted with 'mind-wandering', which Metzinger characterizes as a kind of recurring autonomy loss (2013, p. 1). It is an involuntary, unintentional form of conscious mental behaviour. According to Metzinger (2013, p. 6), the exercise of mental control is significantly less frequent than periods of mind-wandering, amounting to an average of just 9.6 h per day. This would mean that the norm for conscious thought is in fact mind-wandering rather than attentional or cognitive agency. Is this problematic from the point of view of the semantic control needed for the Autonomy View? I do not think there is any reason for the internalist to worry, although there are some issues that she must address.

To start with, note that it is not the case that creation of a new concept requires anything like 9.6 h of controlled conscious thought.[16] As we have seen, in the present metasemantic framework, the creation of a new concept is a matter of a new node being created at a particular location in the conceptual network. As noted above, the conscious behavior responsible for such changes can be quite varied. For example, we may consciously decide to use an existing expression in a new way and this conscious activity will necessitate corresponding changes to the conceptual network. This is a process that may take a matter of seconds (although an agent may wish to spend further time reworking the concept and changes may need to be reinforced over time to prevent it from being forgotten). There is more that the internalist must say here, however. For her approach to be an attractive framework for understanding the creation of new concepts for the purposes of conceptual engineering, it would not be satisfactory to claim merely that new concepts can be created in an appropriately controlled manner. Rather, it must be shown that these new concepts are sufficiently stable over time. Imagine, for example, that during periods of mind-wandering, the conceptual network was substantially altered such that concepts changed their relations to each other at random. If this were the case, there would be little comfort in securing the fact that we can create concepts, for we would be unable to maintain them. This sort of worry may look especially pressing for more holistic conceptual role theories. As noted above, radical holism is thought to entail that any change, however small, to the conceptual network will determine a change in the content of *all* concepts within the network; Jackman (1999) calls this the 'Instability Thesis'. For example, if Jane forms the belief that DOGS ARE FRIENDLY$_J$, this will alter all of her concepts, including those, such as SPANNER$_J$ and DEMOCRACY$_J$, that seem conceptually distant from the concepts employed in the new belief.[17] Some forms of conceptual role theory thus look highly unstable.

[16] It may be that *conscious* thought is not necessary for conceptual engineering. For example, perhaps there can be goal-directed, sup-personal processes that result in conceptual revision. I do not explore this issue here. Rather, I claim that the conscious control described here is sufficient for meeting Cappelen's challenge.

[17] I use an initial in subscript—e.g., DOG$_J$—to indicate that a concept or sentential content is idiosyncratic to the agent (e.g., Jane) whose name begins with that initial.

Fortunately for the holist, whilst there is instability built into the core of her view, it is not of a kind that threatens the Autonomy View. To see this, first note that the instability introduced in the original worry and the instability posited by the holist are of different kinds. The kind of instability driving the original worry is *chaotic*. In a chaotic system, small changes to one part of the system can have unpredictable consequences in other parts of the system. This is the kind of instability involved in Cappelen's Austerity Framework: when a change to the metasemantic base occurs, the effects on meaning cannot be predicted because the algorithm from metasemantic base to meaning (if it exists at all) is too complex for us to grasp. This kind of instability would indeed be problematic for the Autonomy View. However, it is not the kind of instability posited by the holist. For the holist, the network is changing all the time in response to new inputs, but it is changing in systematic and predictable ways. For example, when Jane forms the belief that DOGS ARE FRIENDLY$_{\text{J}}$, it is true that *every* concept in her network changes to a new one. But these changes are not chaotic. Her DOG$_{\text{J}}$ changes to one, DOG*$_{\text{J}}$, that falls under the category FRIENDLY ANIMAL*$_{\text{J}}$. Concepts that represent individual dogs, SPOT$_{\text{J}}$ and CLIFFORD$_{\text{J}}$, will now be classified under the new DOG*$_{\text{J}}$ concept, and their content will change accordingly. Apparently unrelated concepts like SPANNER$_{\text{J}}$ and DEMOCRACY$_{\text{J}}$, will change too. However, again, this will be in predictable ways. For example, assuming that, for Jane, the conceptual relationships that hold between her concepts used to be such that SPANNER$_{\text{J}}$ is not a subcategory of DOG$_{\text{J}}$, the new altered conceptual network instead is such that SPANNER*$_{\text{J}}$ is not a subcategory of DOG*$_{\text{J}}$. This is to say that, although changes are happening all the time, Jane will not suddenly start believing that spanners are dogs. The two concepts, SPANNER$_{\text{J}}$ and SPANNER*$_{\text{J}}$, are very similar before and after the change.[18] A drastic change to the network is certainly possible, but it will always be the result of a specific kind of input to the network, one with predictable consequences.

There is one further issue concerning control that confronts the conceptual role theorist. From the preceding, it may seem that changing one's concepts is a relatively easy process that consists in, for example, deciding on an inferential role for the new concept and, in doing so, fixing its location in the network. However, there is reason to think that, at least in certain cases, deciding on a new inferential role for a concept or expression will not immediately or effortlessly lead to changes in how one employs that concept in reasoning—changing one's actual inferential dispositions, practices, and habits may not be easy at all.[19] One prominent source of this kind of difficulty is the existence of cognitive structures such as stereotypes that are associated with a concept or expression. Stereotypes are intuitive characterisations of a subject matter that represent which features are more salient, central, and important to the agent, as well as intuitive and explanatory connections between them (Camp 2019, p. 19ff). Concepts and stereotypes may be distinct cognitive structures. However, a stereotype can nonetheless exert a strong influence on the inferences that an agent will be disposed to draw—even after she recognises it to be deeply problematic and even when she is strongly motivated to change her dispositions (Devine 1989). Thus, if we cannot easily

[18] Fodor and Lepore (1992) have famously argued that holists do not have workable notion of conceptual similarity. For a response to this objection for the conceptual role theorist, see Pollock (2020).

[19] Thank you to an anonymous reviewer for raising this issue.

alter the stereotypes associated with an expression, it may be difficult to create a new concept with our desired inferential profile.

It is worth noting that, even if it could be demonstrated that stereotypes are never an obstacle to conceptual engineering per se, there is something strange about a conceptual engineer who is interested in engineering a concept with a view to some ethical benefit, but who does not care about changing any problematic stereotypes that accompany it. For example, imagine a conceptual engineering project that seeks to ameliorate the concept MARRIAGE such that it is no longer restricted to heterosexual partnerships, but that does nothing to address a stereotype that represents heterosexual marriages as a more central example of the phenomenon. This is clearly still problematic and at odds with the engineer's aim of promoting equality for the LGBT+ community. It seems, then, that conceptual engineers ought to be concerned with changing stereotypes, such that excluding them from engineering projects would be inappropriate.

With this in mind, the approach that I think the conceptual role theorist should take is to maintain that stereotypes are not an obstacle to concept *creation*, but allow that they may be an obstacle to the *implementation* of the new concept—even within the individual who created it. This response requires a complication of the view as introduced in Sect. 4.1. On this new approach, it is possible to create a concept without implementing it, even in oneself. What does this mean? The idea is that, whilst creating or acquiring a new concept requires merely that one *entertain* this concept, implementing this concept requires that one *endorse* it (Pollock 2019). In merely entertaining a concept, one recognises which inferences the concept licenses (just as one may entertain or understand a stereotype without endorsing or implementing it in cognition).[20] In endorsing a concept, one gains the disposition to actually apply the concept, and to draw the inferences that it licenses. I think that this distinction is a natural one for conceptual engineers to employ.[21] To see why, consider that engineers who are engaged in ameliorative projects will often find themselves with a rich understanding of *both* the original, problematic concept, and the new, ameliorated concept that they wish to replace it with. For example, those who are interested in the amelioration of MARRIAGE may have extensive knowledge concerning the original concept and its current and historical role in civil society (see, e.g., Calhoun 2002); this understanding is extremely useful for both developing and motivating the ameliorated concept. Such a conceptual engineer will possess both the deficient and the ameliorated concept, but she will only *endorse* the latter. This engineer will recognise which inferences the original concept licenses, without being disposed to draw these inferences—indeed, whilst maintaining that these inferences should be rejected.[22] Returning to the present problem, the suggestion is that an individual may succeed in creating an ameliorated concept, but to implement this concept, even within herself, she must come to endorse it, and *this*

[20] This use of 'endorse' is slightly different to the use of this terminology in Camp's (2019) explanation of stereotypes and characterisations.

[21] It is worth noting that the distinction is plausible independent of the present debate. Rabin (2020) offers a number of arguments for maintaining the distinction. For example, the distinction allows us to handle cases in which an agent understands a slur but does not endorse it. Rabin's view is that mastering a concept involves recognising the rules governing its application.

[22] The original concept may even remain in the network if it is useful (e.g., because it is of historical interest), although it will likely become associated with a new word-form to avoid confusion.

may require that she alters her dispositions to employ certain stereotypes and draw certain inferences. To the extent that these dispositions are resistant to change, this means that, although the creation of an ameliorated concept may remain relatively easy, implementation, even within the engineer herself, may be significantly more difficult.

I have argued that the conceptual role theorist can recognise the relevance of stereotypes to conceptual engineering whilst maintaining that concept creation is relatively easy. However, what if one does not wish to maintain the distinction between entertaining and endorsing a concept? I think that, even if one wishes to maintain that stereotypes do directly impede the creation or revision of concepts, the internalist still can accommodate this whilst succeeding in meeting Cappelen's challenge. Stereotypes are certainly difficult to change; however, there is a wealth of research in social cognition that suggests that they are malleable and can be altered in predictable ways. There is evidence both that individuals are capable of counteracting or preventing the automatic activation of biases (Devine 1989; Moskowitz et al. 1999), and that individuals can identify and change stereotypes over time (Blair 2002; Lenton et al. 2009). In relation to Cappelen's challenge, it is important to note that, although it is difficult to effect such changes, conceding this is very different from conceding that this sort of change is chaotic, unsystematic, inscrutable, or beyond our control. Thus, even if the existence of things like stereotypes forces us to accept that concept creation is difficult, the internalist can still reject both Inscrutable and Lack of Control.

Thus far, I have identified some states to which we have epistemic access and over which we have a reasonable degree of control. Why think that concepts supervene on *those* inner states? In the next section, I take up the second element of Cappelen's challenge.

4.2.2 Concepts supervene on those inner states

Cappelen, like many externalists, will treat the factors appealed to in conceptual role theory as 'metasemantic superstructure'. Metasemantic superstructure comprises the attitudes people have regarding concepts or word meanings—what they believe, hope, desire, etc. these concepts or meanings to be. Depending on the form of externalism that you endorse, metasemantic superstructure will have little effect on the meaning facts (as noted above, Cappelen allows that it plays a minor role in content determination). What grounds do we have to think that the conceptual network is the total metasemantic supervenience base for content rather than merely consisting in a network of beliefs about content?

The answer to this aspect of Cappelen's challenge is simple. Conceptual role theory just is the view that the conceptual network determines concepts. Dialectically, the approach one usually takes when arguing for one's theory of concepts is to set out what one thinks the desiderata are for a good theory of concepts and then show how one's theory meets these desiderata (and, if possible, how competing theories fail).[23] The desiderata here are things like explanations of action, rationality, communication, disagreement, categorisation, learning, and, possibly, conceptual engineering. Of course,

[23] For examples of this approach, see Block (1986) and Prinz (2002).

for her view to be ultimately defensible, a conceptual role theorist must demonstrate two things: firstly, that her view does indeed satisfy her preferred desiderata and, secondly, that these desiderata are appropriate desiderata for a theory of concepts. However, the project of this paper is not to comprehensively defend conceptual role approaches to content. Rather, the aim is to explain how the view could meet Cappelen's challenge to internalism. If internalist conceptual role theories turn out to be all-things-considered bad theories of concepts, then they should be rejected. But, I claim, if true, they provide a promising framework for understanding conceptual engineering. That they can do this is, I think, one mark in favour of this kind of view. Let us turn to the final element of Cappelen's challenge.

4.2.3 The determination relation from supervenience base to content is scrutable and within our control

The last part of Cappelen's challenge is for the internalist to defend the claim that the *determination relation* from metasemantic supervenience base to content is itself scrutable and within our control. There are two parts to this challenge. The first is epistemic: if we don't know the algorithm that takes us from metasemantic base to content then, even if we do have epistemic access to the metasemantic base (as I have argued that we do), we still won't be able to maintain the Autonomy View because we will not be able to discern how changes to this base effect changes in content. The second part of the challenge concerns control: Cappelen suggests that the algorithm itself might change over time in ways that we cannot control. In response, I will argue for two things: (a) internalists can indeed maintain that we have epistemic access to the determination relation, and (b) we do not have control over this relation itself, but we do not need it.

As noted above, Cappelen presents internalism as a supervenience claim: meaning supervenes solely on factors internal to the agent. In response to the epistemic part of Cappelen's challenge, first note that, whilst it is likely true that all internalists accept the supervenience claim, many internalists make stronger claims than this regarding the relationship between supervenience base and content. Cappelen never denied this, of course: his point was that, thus far, no internalist had made a case for the claim that we have access to the supervenience relation itself such that deliberate conceptual engineering is possible. In what follows, I identify some versions of conceptual role theory that can claim this.

The most straightforward way for the conceptual role theorist to respond to the epistemic issue is to claim that concepts are just identical with their metasemantic base. On this approach, it is easy to track how changes to the metasemantic base relate to changes in content because changes to this base just *are* changes in content. There are different options the conceptual role theorist might choose here. She could claim a view analogous to 'realizer' functionalism in philosophy of mind (Lewis 1966). Realizer functionalism is the view that a given mental phenomenon, such as pain, is identical with the physical entities that realize the functional role characteristic of the phenomenon—C-fibres, for example. This kind of approach, applied to concepts, is the view that concepts are identical with certain physical elements in the brain—they are the 'occupants' of locations in the network. Any internalist that is willing to commit

to a mind-brain identity theory with respect to concepts can make the same kind of move in response to Cappelen. My preference is for an identity theory of some form[24]; however, this isn't the only option available to the conceptual role theorist. She could instead opt for an account analogous to role functionalism (Putnam 1967). Role functionalism is the view that a given mental phenomenon is identical with a role state. On this view, pain (for example) is a higher-order state or property: it is the state of being in a state that plays the pain role. Two individuals are in pain if they are in a state that plays the pain role, regardless of the physical properties of this realizer state. The analogous view, applied to concepts, would be that concepts are identical with conceptual roles, rather than their realizers. Whichever option one picks, the general point is that Cappelen's worry only arises if we appeal to *mere* supervenience in defining a particular internalist position. This is what leaves room for meaning and concepts to supervene on their metasemantic base in surprising ways. If concepts are identical with the base that individuates them, however, the worry does not arise.

What of the second aspect of Cappelen's challenge? On the approach I am suggesting, we do not have control over the relationship between metasemantic base and content in the following sense: we cannot bring it about that content supervenes on different factors through sheer force of will (or any other means for that matter). However, I do not think this is a problem for the internalist. As explained above, the view I defend is that concepts are identical with their metasemantic base; it is a feature of this view that this relationship is not something that can change over time. As such, given that we do have control over the metasemantic base, and we do know how changes to this base will result in stable and predictable changes to our concepts, we have sufficient control to maintain the Autonomy View.

This approach will not be attractive to all—not even all internalists about content individuation. When it comes to the ontology of concepts, there is a division amongst those who think that concepts are mental representations (or similar) and those who think that contents are abstract objects (e.g., Peacocke 1992; Rey 1994). Any internalist who occupies this latter camp will not be able to adopt the response to Cappelen's challenge that I am suggesting here: one cannot identify a concept with its metasemantic base if the former, but not the latter, is an abstract object. Having said that, a great many contemporary internalist theories of concepts do take concepts to be mental representations and, as such, my approach should have wide applicability. For example, amongst theories of concepts are those that identify concepts with prototypes (Rosch and Mervis 1975), exemplars (Estes 1994), proxytypes (Prinz 2002), and simulators (Barsalou 1999).[25] Each of these theories can treat concepts as psychologically real entities and can thus opt for the same kind of response to Cappelen that I suggest for the conceptual role theorist.

[24] Such a view must deal with the objection from multiple realizability. This objection has historically been extremely influential; however, more recently, many authors have argued against multiple realizability with respect to certain kinds of mental state, and also against the claim that the phenomenon poses a problem for identity theories (see e.g., Shagrir 1998; Polger 2009; Shapiro 2000; Bickle 1998; Zangwill 1992; Bechtel and Mundale 1999; Couch 2004; Miłkowski 2016).

[25] See Laurence and Margolis (1999) for an overview. It is, perhaps, more common for theorists to treat word meanings as abstract objects (e.g., Pagin 2006; Katz 1981); however, there are also prominent defenders of the view that meanings or languages are properties of the brain (e.g., Chomsky 1986).

In this last strand of my response to Cappelen's challenge, I think we see an advantage for an internalist theory of concepts. On externalist views, it would typically be implausible to identify a concept with its metasemantic base. For example, even supposing that it is plausible to claim that concepts are individuated by complex patterns of use in a community, it seems less plausible that a concept *is* a pattern of use in a community. It is not clear how patterns of use in a community could be compositional, or employed by an individual as categorisation devices, or be expressed in communication. A similar thing goes for causal chains, or nomic relations to environmental properties. Externalists are, I think, more likely to claim that, although the *vehicles* of content are physical and internal, these contents themselves are abstract objects.[26] For this reason, they do not have available to them the same kind of defence against Cappelen's concerns regarding supervenience.

5 Conclusion

Cappelen challenged the internalist to argue for three claims: (a) there are inner states that are scrutable and within our control; (b) meaning supervenes on these inner states; and (c) the determination relation from supervenience base to meaning is scrutable and within our control. In this paper, I have argued that internalist conceptual role theories can meet each strand of Cappelen's challenge. On the view that I propose, concepts are to be identified with the realizers of conceptual roles. This view can maintain that concepts are entities that we have a suitable degree of control over, and epistemic access to. The internalist conceptual role theorist enjoys two advantages over at least some externalist (and perhaps also some internalist) competitors. Firstly, she can reject both Inscrutable and Lack of Control and thus maintain the Autonomy View: it is possible to analyse and improve our concepts in a purposive and controlled way. This is a *pro tanto* reason to prefer (certain kinds of) internalism *if* one thinks that it is a desideratum on a theory of concepts that it can claim that this kind of engineering is possible. Secondly, the internalist view that I defend here can offer a distinctive response to Cappelen's worries about supervenience that is not available even to less radical forms of externalism that might also wish to reject the Austerity Framework.

Acknowledgements Open Access funding provided by University of Oslo (incl Oslo University Hospital). Many thanks to Mirela Fuš, Sigurd Jorem, Mark Pinder, Delia Belleri, Manuel Gustavo Isaac, and two anonymous reviewers for this journal for their very helpful comments. Thank you also to audiences at the University of Oslo, University of Toronto, and the Czech Academy of Sciences. This research was funded by the Research Council of Norway (250654).

[26] Externalists (and internalists) may also claim that referents are part of the physical world. To claim this, however, is not to claim that a representational device is identical with its metasemantic base and, thus, this will not help in meeting Cappelen's challenge.

References

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.
Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science, 66,* 175–207.
Bickle, J. (1998). *Psychoneural reduction: The new wave*. Cambridge, MA: MIT Press.
Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6,* 242–261.
Block, N. (1986). Advertisement for a semantics for psychology. *Midwest Studies in Philosophy, 10,* 615–678.
Boghossian, P. (1994). The transparency of mental content. *Philosophical Perspectives, 8,* 33–50.
Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy, 4,* 73–121.
Burge, T. (1988). Individualism and self-knowledge. *Journal of Philosophy, 85,* 649–663.
Burgess, A., & Plunkett, D. (2013). Conceptual ethics I. *Philosophy Compass, 8,* 1091–1101.
Calhoun, C. (2002). *Feminism, the family, and the politics of the closet: Lesbian and gay displacement*. Oxford: Oxford University Press.
Camp, E. (2019). Perspectives and frames in pursuit of ultimate understanding. In S. R. Grimm (Ed.), *Varieties of understanding: New perspectives from philosophy, psychology, and theology*. Oxford: Oxford University Press.
Cappelen, H. (2018). *Fixing language*. Oxford: Oxford University Press.
Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. New York: Praeger.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58,* 7–19.
Couch, M. (2004). Discussion: A defense of Bechtel and Mundale. *Philosophy of Science, 71,* 198–204.
Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.
Eklund, M. (2014). Replacing truth? In A. Burgess & B. Sherman (Eds.), *Metasemantics: New essays on the foundations of meaning* (pp. 293–310). Oxford: Oxford University Press.
Estes, W. K. (1994). *Classification and cognition*. Oxford: Oxford University Press.
Field, H. (1977). Logic, meaning and conceptual role. *Journal of Philosophy, 69,* 379–408.
Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
Fodor, J., & Lepore, E. (1992). *Holism: A shopper's guide*. Oxford: Blackwell.
Harman, G. (1987). (Non-solipsistic) conceptual role semantics. In E. Lepore (Ed.), *New directions in semantics*. London: Academic Press.
Haslanger, S. (2000). Gender and race: What are they? What do we want them to be? *Nous, 34,* 31–55.
Haslanger, S. (2006). Philosophical analysis and social kinds: What good are our intuitions? *Proceedings of the Aristotelian Society, 80,* 89–118.
Hudson, R. (2007). *Language networks: The new word grammar*. Oxford: Oxford University Press.
Jackman, H. (1999). Moderate holism and the instability thesis. *American Philosophical Quarterly, 36,* 361–369.
Katz, J. (1981). *Language and other abstract objects*. Totowa, NJ: Rowman and Littlefield.
Koch, S. (2018). The externalist challenge to conceptual engineering. *Synthese*. https://doi.org/10.1007/s11229-018-02007-6.
Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 3–81). Cambridge, MA: MIT Press.
Lenton, A. P., Bruder, M., & Sedikides, C. (2009). A meta-analysis on the malleability of automatic gender stereotypes. *Psychology of Women Quarterly, 33,* 183–196.
Lewis, D. (1966). An argument for the identity theory. *Journal of Philosophy, 63,* 17–25.
Metzinger, T. (2013). The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers in Psychology, 4,* 931.

Miłkowski, M. (2016). Computation and multiple realizability. In V. C. Mueller (Ed.), *Fundamental issues of artificial intelligence*. Berlin: Springer.
Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77,* 167–184.
Nado, J. (2019). Conceptual engineering, truth, and efficacy. *Synthese*. https://doi.org/10.1007/s11229-019-02096-x.
Pagin, P. (2006). Meaning holism. In E. Lepore & B. Smith (Eds.), *Handbook of philosophy of language*. Oxford: OUP.
Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.
Pinder, M. (2019). Conceptual engineering, metasemantic externalism and speaker-meaning. *Mind*. https://doi.org/10.1093/mind/fzz069.
Polger, T. (2009). Evaluating the evidence for multiple realization. *Synthese, 167,* 457–472.
Pollock, J. (2019). Conceptual engineering and semantic deference. *Studia Philosophica Estonica, 12,* 81–98.
Pollock, J. (2020). Holism, conceptual role, and conceptual similarity'. *Philosophical Psychology*. https://doi.org/10.1080/09515089.2020.1729973.
Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: MIT Press.
Putnam, H. (1967). The nature of mental states. In H. Putnam (Ed.), *Language, mind & knowledge*. Minneapolis: University of Minnesota Press.
Putnam, H. (1975). The meaning of "meaning". In H. Putnam (Ed.), *Language, mind & knowledge*. Minneapolis: University of Minnesota Press.
Rabin, G. (2020). Toward a theory of concept mastery: The recognition view. *Erkenntnis, 85,* 627–648.
Rapaport, W. (2002). Holism, conceptual-role semantics, and syntactic semantics. *Minds and Machines, 12,* 3–59.
Recanati, F. (2012). *Mental files*. Oxford: Oxford University Press.
Rey, G. (1994). Concepts. In S. Guttenplan (Ed.), *A companion to the philosophy of mind*. Cambridge, MA: Blackwell.
Richardson-Self, L. (2015). *Justifying same-sex marriage: A philosophical investigation*. London: Rowman & Littlefield.
Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.
Shagrir, O. (1998). Multiple realization, computation and the taxonomy of psychological states. *Synthese, 114,* 445–461.
Shapiro, L. (2000). Multiple realizations. *Journal of Philosophy, 97,* 635–654.
Wikforss, A. (2015). The insignificance of transparency. In S. Goldberg (Ed.), *Externalism, self-knowledge, and skepticism* (p. 2015). Cambridge: Cambridge University Press.
Williamson, T. (1994). *Vagueness*. London: Routledge.
Zangwill, N. (1992). Variable reduction not proven. *Philosophical Quarterly, 42,* 214–218.