Canadian Journal of Philosophy (2020), 50: 5, 594–605
doi:10.1017/can.2020.8

SYMPOSIUM
ON HERMAN CAPPELEN'S *FIXING LANGUAGE*

Conceptual Engineering, Topics, Metasemantics, and Lack of Control

Responses to Sarah Sawyer, Laura Schroeter, François Schroeter, and Tim Sundell

Herman Cappelen

University of Oslo, Oslo, Norway
Email: herman.cappelen@gmail.com

Conceptual engineering is now a central topic in contemporary philosophy. Just four to five years ago it wasn't. People were then engaged in the engineering of various philosophical concepts (in various subdisciplines), but typically not self-consciously so. Qua philosophical method, conceptual engineering was underexplored, often ignored, and poorly understood. In my lifetime, I have never seen interest in a philosophical topic grow with such explosive intensity. The sociology behind this is fascinating and no doubt immensely complex (and an excellent case study for those interested in the dynamics of academic disciplines). That topic, however, will have to wait for another occasion. Suffice it to say that if *Fixing Language* (*FL*) contributed even a little bit to this change of focus in philosophical methodology, it would have achieved one of its central goals. In that connection, it is encouraging that the papers in this symposium are in fundamental agreement about the significance and centrality of conceptual engineering to philosophy. That said, the goal of *FL* was not *only* to advocate for a topic, but also to defend a particular approach to it: the Austerity Framework. These replies have helped me see clearer the limitations of that view and points where my presentation was suboptimal. The responses below are in part a reconstruction of what I had in mind while writing the book and in part an effort to clarify and improve my view. I'm grateful to the symposiasts for helping me get a better grip on these very hard issues.

Reply to Tim Sundell

Sundell's Objection to the Contestation Theory of Topic Continuity

In *FL*, I claim that an expression's intension can change (what "*F*" denotes can change), but the topic being talked about by those using "*F*" can be constant: topic continuity can be preserved under semantic change. There can, I argue, be continuity in *F*-inquiry (and agreement and disagreement about *F*s expressed using "*F*") despite the fact that what "*F*" picks out changes. Sundell wants to know what underwrites topic continuity in such cases and argues that *FL* fails to give an adequate answer.

Sundells's formulation of the problem goes like this: Suppose we theorists know that speakers *A* and *B* are talking about different kinds of things when they use "*F*" (because we know that the things picked by "*F*"—the intension—has changed.) How, then, can we, *qua theorists* (this part will be important), justify the assumption that when *A* says "*Fs* are not *G*" and *B* says "*Fs* are *G*," they really disagree? After all, we know that they are talking about different kinds of things, so we have a theoretical reason to resist the disagreement judgement.

Sundell grants that in these cases, it will often be true to utter the English sentence, "They are both talking about *Fs*," but, he objects, that observation alone is not sufficiently explanatory. We theorists

want to know by virtue of what there is an underlying continuity of topic that makes true the utterance. To be offered a piece of trivially true nontheoretical discourse ("They both talked about *F*") doesn't respond to the theoretical worry. My reply, according to Sundell, is a kind of luddite refusal to give a theory—a simplistic repetition of the ordinary language sentence "they are both talking about *F*," but with no theoretical account of what makes that claim true or how it is responsive to the theoretical concern.

Sundell has an illustration of what would constitute a candidate answer: what is preserved is a functional role. That functional role can be in place even when the intension changes. Sundell (2011, 2017), Plunkett and Sundell (2013), Haslanger (2000), Thomasson (2019), and many others defend this kind of view.

A Three-Part Defense of the Contestation Theory of Topic Continuity

1. A general theory of how to find unity in diversity

I suspect that one reason Sundell thinks I fail to give a real theoretical account of topic continuity is because he is committed to the view that what creates stability (in topic) is some kind of invariant feature (e.g., a function) that is present throughout change. That, however, is hardly ever the answer to a question about stability through change. This passage from Timothy Williamson (2007, 123) explains why:

> What binds together different events into the history of a single complex object, whether it be a stone, a tree, a table, a person, a society, a tradition, or a word? In brief, what makes unity out of diversity? Rarely is the answer to such questions the mutual similarity of the constituents. Almost never is it some invariant feature, shared by all the constituents and somehow prior to the complex whole itself—an indivisible soul or bare particular. Rather it is the complex interrelations of the constituents, above all, their causal interrelations.

For Sundell, it is the functional role that creates unity (topic continuity) in diversity (of changing intensions). In *FL* (chapter 16), I argue that this is wrongheaded not just in detail (i.e., it is not just a matter of finding the right kind of function or object), but in principle. Unity is the result of historical continuity of a complex interaction between intensions, expressions, and language use. It is not generated by a *something* that is present throughout the process of revision. If Sundell thinks a "theoretical" account of topic continuity has to explain continuity by such a presence, then he's asking me for something he'll never find.

2. My theory of topic continuity

What I just said is not a refusal to engage in theorizing. It is, instead, to base the answer on a theory: the theory that unity isn't based on an invariant shared element, but instead on historical continuity and interconnectedness. My theory of topic continuity so understood has many parts and I'll briefly mention five of them:

1. *Causal-Interrelations*: Central to topic continuity is causal interrelations between the varying stages. The focus on causal continuity is familiar from other domains.
2. *Genealogies*: This is not to deny that there are a great deal of interesting things that can be said about particular evolutionary histories. This is what genealogies will illuminate when they are done right. They are done right when they track a philosophically interesting object; on my view, that would be a topic that persists over time (even as the associated expressions undergo semantic change). It is crucial that this be sharply distinguished from what are simply changes in beliefs, practices, or theories. That distinction, however, is hardly ever made; it's all mixed together into a rather confused dog's dinner. As a result, existing genealogies are hardly every useful (from the point of view of conceptual engineering), but in principle they would be. As I

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

see it, there's an entire research field of *conceptual engineering informed genealogy* waiting to be developed.[1]

3. *Checklists*: We might, as I point out in *FL*, also find useful generalizations about topic continuity (2018, 120–21). That's the point of the discussion of Railton's checklist in that part of the book. I like Railton's list because he explicitly denies that these are necessary or sufficient conditions.
4. *Contestation as a Constitutive Component of Topic-Continuity*: I propose that what constitutes topic continuity is in part determined contextually by conversation partners. In conversation *C* there can be disagreement about whether *A* and *B* are talking about the same topic, and part of the answer to that question might depend on what happens in *C*. Maybe this is a form of antirealism about topic continuity; for more on that, see the reply to Schroeter and Schroeter below.
5. *Interpreter relativity*: I argue that topic continuity is relative to interpreters' interests. This interpreter-relativity of topic continuity is continuous with the even more radical claim that I have argued for elsewhere—that what speakers say (and assert) is relative to interpreters (see my 2008, according to which speaker *A* by uttering a sentence *S* in context *C* can have said one thing relative to interpreter *I1* and something else relative to interpreter *I2*).

Think of my overall theory of topics as having three parts: (i) 1–5, (ii) the arguments against reification of topics (in chapter 16 of *FL*), (iii) the rejection of the assumption that unity of topic has to be accounted for by the presence of an invariant feature.

Sundell asks for "… a theoretical account of how the relevant inquiries could be considered continuous" (2020, 9)[2] and I've given him that. It's a different kind of theory from what he has in mind. It's not an effort to provide reductive necessary and sufficient conditions, and it doesn't explain unity by appeal to the presence of an invariant feature. If Sundell doesn't think it's worthy of the label "theory," then he is working with a mistaken view of what constitutes a theory.

Sundell also asks for a theoretical account of how the disagreements between speakers at different stages of a process of amelioration could be considered "substantive" (9). Using the example from above, the speakers are disagreeing over what families are. A disagreement about that is substantive and important because it has massive implications for many aspects of personal and social life. Their different views about families can affect their actions, goals, and plans in all kinds of ways. If Sundell doesn't want to call that "substantive," then he's wrong about what it takes to be "substantive."

There is a lingering concern and it explains why Sundell probably will be unmoved by what I just said. Here is how Tristram McPerson (pers. comm.) helpfully puts it: "I am tempted to accept the whole account as a plausible theory of our folk phrase 'same topic,' and suggest that this account brings out exactly the substantive uninterestingness of this folk concept for theoretical purposes, and the need to engineer something in the vicinity more theoretically interesting." Why do we need something more theoretically interesting? Here is how I think someone pushing Sundell's objection would answer:

> We theorists know that when two speakers use "family" with different semantic values, then what *A* says (preamelioration) by uttering "Families are *G*" and what *B* says (postamelioration) by uttering "Families are not *G*" can both be true. We therefore know that there's a sense of "substantive disagreement" that's not captured by the ordinary notion of disagreement (the one that's focused on topics). This more interesting sense is one according to which *A* and *B* don't disagree (despite it being true that *A* says *that Families are* G and B *says that families are not* G).

[1]This will be a big field and the part of it that's relevant for the reply to Sundell is that these genealogies can help us understand topic continuity by detailed description of examples of such continuity (and discontinuity).

[2]In the "Reply to Tim Sundell" section, unless otherwise stated, all folios refer to Sundell's 2020 article, "Changing the Subject."

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

This version of the objection still fails to take my theory fully onboard. On my view, what *A* and *B* say (i.e, *that Families are* Gs and *that Families are not* Gs) should not be conflated with the semantic value of the sentences they use to say it (i.e., the sentences "Families are *G*" and "Families are not *G*"). What they say something about is about families (construed as a topic) and they disagree about that.

In response to this, Sundell's objection can be extended as follows: *Okay, that just shows we need to do even more engineering: we need to engineer a notion of "saying" so that it is focused on the semantic content and we need to use this reengineered notion to develop a better notion of "substantive" disagreement.* At this point, I'll get cautiously concessive: Sure, you could do that, but you might not need to do much engineering. You just need to get in the right kind of interpreter context. Recall, according to 5 above, that what speakers say varies between contexts of interpretation and there could be interpreters (maybe philosophers or semanticists) who obsessively focus on semantic contents. Nothing I've said rules that out. However, if you find yourself in that kind of context, you shouldn't assume that you're in some kind of intellectually superior interpretative context. You just care about a level of content that most others don't care so much about. To deflate the temptation to privilege this kind of context, it will help to reflect on the fact that very many people in other contexts of interpretation care deeply about disagreements (they sometimes lead to wars) and what they care about (and fight their wars over) isn't some hard-to-pin-down content that a few philosophers want to focus on in a few theoretical contexts. So to capture what matters to most people, you have to escape the quirky restrictions of idiosyncratic interpreter contexts.

3. Against functional role as the source of topic continuity

Finally, a brief comment on why I don't think an appeal to the functional role of a concept can explain topic continuity. I'll focus on the least controversial argument I have against the view[3]: one of the elements up for revision when doing conceptual engineering is that of functional role itself. It is not sacrosanct and is as much up for revision as any other component of the evolutionary process. Put in the terminology of *FL*, one of the reasons why one might want to change the intension of an expression is *because* one wants that expression's functional role to change. Topic change doesn't line up with change in functional role. The functional role of, say, "family" (or the concept it expresses), can change a little bit, but the topic—family—can be continuous through that change. This is evidenced by (not constituted by) the acceptability of certain kinds of disquotational reports and agreement and disagreement judgements (i.e., the sorts of evidence that Sundell appeals to in his own account).

Reply to Laura Schroeter and François Schroeter

The Metasemantic Foundations of Conceptual Engineering

FL is in part an elaboration of the metasemantic foundations of conceptual engineering and the implications of these. The book doesn't argue for a new metasemantic theory, but instead builds on familiar externalist theses that, while not uncontroversial, are endorsed by quite a few philosophers. Since the book is not about metasemantics, there are many important issues in that vicinity that I don't address. Doing so would involve raising questions the answer to which have little or no direct bearing on the other claims I make in *FL*. Schroeter and Schroeter, however, argue that I have to say more, and they give me some options for what to say. The structure of my reply goes like this. First, I point out that, for my purposes, I don't need to say more than I do. Second, I don't like any of the options they offer me. Finally, I outline what I would say if I were to say more.

[3]The more controversial objection is that I don't think expressions have functions of the kinds that many authors appeal to. If there are no such functions, then they can't explain topic continuity. A defense of that claim would take us too far afield but can be found in *FL* chapter 16.

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

Schroeter and Schroeter's Objection

Initially, Schroeter and Schroeter summarize an important part of my view in a nice way: "(i) The *inputs into interpretation* are not (and cannot be) fully known." "(ii) The *interpretation function* is not (and cannot be) fully known." This isn't exactly how I put it, but it's a very useful articulation of a core idea. Schroeter and Schroeter then go on to ask: "… what exactly does it mean to say a term's reference (or a change in its reference) is inscrutable? Cappelen doesn't say" (2020, 4).[4] They then go on to offer me two options which they call "Unknown Boundaries" and "Cluelessness." They then argue that the latter thesis is false (and doesn't follow the kind of moderate version of externalism that I endorse), while the former thesis is too weak and irrelevant to conceptual engineering.

What I Say about Metasemantics and Why I Don't Say More

The externalist assumptions in *FL* were introduced to raise a cluster of concerns, most of them having to do with our control (and lack thereof) over the metasemantic foundations of the languages we speak. Here is Schroeter and Schroeter's description of what I have in mind: "Because we can't know which changes in our linguistic practices would trigger a change in reference, any effort to systematically plan and implement changes in the reference of our words is bound to be highly unreliable" (3). The key question for me, as I was writing the book, was whether I said enough *for my purposes*. So one crucial question then is, do (i) and (ii) suffice to establish what I call "Lack of Control"?

I agree with Schroeter and Schroeter that Lack of Control doesn't follow logically from (i) and (ii). However, if we combine (i) and (ii) with some rather minimal assumptions, then Lack of Control becomes hard to resist:

Additional Assumptions:

a. Complicated facts about history (dubbings and communicative chains) play an important role in fixing intensions. It is indisputable that we have no way to change that history and also that much of it is unknowable.
b. Use patterns over time play an important role in fixing intensions and "the use patterns of an English word" refers to literally billions of speech events over very long periods of time. Again, it is indisputable that none of us have any significant degree of control over the sum of use patterns for even one word in one language.
c. Even if we knew all the grounding facts for intensions of particular expressions (which we don't), we don't know how they add up to a semantic value: we don't know the supervenience relation.

It is ((i)–(ii)) + (a–c) I rely on as motivation for Lack of Control. I can say all of that without taking a stand on Cluelessness and Unknown Boundaries. In sum: *I think Schroeter and Schroeter are asking for commitments where agnosticism serves me better.*

Disquotational Knowledge versus Cluelessness and Unknown Boundaries

That said, Schroeter and Schroeter raise important issues in metasemantics that I have views about. First, I reject Cluelessness[5] because, as Schroeter and Schroeter note and discuss (9ff), I think we know disquotationally *exactly* what our words denote (e.g., we know that "salmon" denotes all and only salmons and that "child" denotes all and only children). Moreover, I think that we care about

[4]In the "Reply to Laura Schroeter and François Schroeter" section, unless otherwise stated, all folios refer to their 2020 article, "Inscrutability and Its Discontents."

[5]I reject Unknown Boundaries too, but given limitations of space, I will focus on Cluelessness.

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

that disquotational knowledge (for a discussion of this see page 83 of *FL*). On the face of it, that commitment is incompatible with Cluelessness (so it is strange to even consider attributing that view to me).[6]

Can All Our Nondisquotational Beliefs about a Referent Be wrong?

What Schroeter and Schroeter have in mind is a version of Cluelessness that is compatible with such disquotational knowledge. One way to capture that would be to just exclude disquotational knowledge explicitly from their formulation of Cluelessness.[7] If we do that, we get to a point of substantive disagreement. So understood, Schroeter and Schroeter's claim becomes:

> It would be a mistake to take the mere possibility of error to support Cluelessness: even if any of our [nondisquotational] assumptions could turn out to be wrong, it does not follow that all of our [nondisquotational] assumptions could be wrong at once. (5; my additions in square brackets)

They are right, of course, that Cluelessness (so understood) doesn't follow from the mere *possibility* of error. That said, I think reflection on particular cases lends support to the view that, in some contexts, given certain historical and contextual factors, someone can denote *o* using "*N*" even when all the speaker's nondisquotational (and nonmetalinguistic) beliefs about *o* are wrong. I have in mind cases where speaker, *A*, has very few beliefs and those are all wrong: *A* thinks "Nancy" refers to a person when it instead denotes a mathematical object, say a Lie Group.[8] Given the right historical connections and deferential intentions, *A* can use "Nancy" to denote Nancy (i.e., the Lie Group) even when *A* doesn't know Lie Groups exist. If *A* utters "I think Nancy is a person," *A* has said something true because *A* has the belief that Nancy is a person (and that belief is false since Nancy is a Lie Group). This is, when fully spelled out, a case where all the nondisquotational (and nonmetalinguistic) beliefs are false.

Why We Care about Semantic Content: The Primacy of Sayings

Schroeter and Schroeter don't discuss particular cases like this, but they have a general argument for why there cannot be any: it undermines what they describe as "the theoretical points of" attributing semantic contents to terms.[9] Semantic theorizing, according to Schroeter and Schroeter, plays two types of theoretical roles: (a) *explanatory roles*: e.g., predicting and explaining individual speakers' behavior, or explaining the propagation of a term within a community; (b) *normative roles*: e.g., explaining which uses of a term are rational, semantically correct, or true.

I disagree with this characterization of semantics: there is no unique set of theoretical roles that semantic contents (or theories) play. This varies considerably between theories. In some, the aim is to give a compositional semantics that builds on compositional syntactic rules. Some think the goal

[6]It is incompatible with the claim that "given our actual cognitive limitations, none of our assumptions about what's represented by our words can be known to be true." it is also incompatible with Unknown Boundaries, but, as I said, I won't discuss that further here.

[7]Cluelessness would have to be revised as follows: *Ca. Given our actual cognitive limitations, none of our nondisquotational assumptions about what's represented by our words can be known to be true. Cb. In particular, our methodological assumptions about how to get closer to the truth about what's represented cannot be trusted to be reliable.*

[8]See Kripke (1980, 116n). Kripke doesn't rule this out as a case of reference (and emphasizes that if it isn't, then the reason isn't a lack of true beliefs).

[9]Reminder of point above: suppose Schroeter and Schroeter are right that there is some small cluster, *C*, of true beliefs about *o* that a speaker must have in order to use "*N*" to denote *o*. That point alone is orthogonal to the central arguments in *FL*. This is especially so because (i) *C* is not the same for all speakers—it is not a common creed (see Schroeter and Schroeter, 2), (ii) *C* isn't sufficient to fix reference in the speech community, and (iii) the existence of *C* doesn't give insight into how we change reference in the community—and, in particular, doesn't give us control over that process.

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

of semantics is to model (at a certain level of abstraction) a structure found in speakers' brains. For others, the aim is to find an anchoring point for reference as speakers' beliefs and theories changes. I use semantic contents to anchor topics (for more on this, see the discussion of Chalmers and Eklund in *FL* chapter 18, and Cappelen 2019). However, my general point is independent of any of these particular views: It's a mistake to fix on theoretical roles and then use those to rule out a view of semantics. That puts the cart before the horse.

There is another equally fundamental disagreement between us in the vicinity: many of the roles that Schroeter and Schroeter assign to *semantic content*, I assign to *speech act content. What speakers say* (by uttering a sentence) is fundamentally different from *what they semantically express.* On my view, it is what speakers say that plays a central explanatory and normative role. Speech act content is the primary explanatory component, not semantic content. Assigning communicative primacy to what speakers say (and not to semantic content) has consequences that are relevant to some of Schroeter and Schroeter's more specific objections. Consider the following claim from Schroeter and Schroeter: "If reference is inscrutable, an accurate answer to "what is *x*?" questions is forever beyond the cognitive reach of both competent speakers and theorists alike" (11). Not true. What the speaker asks when he or she utters "What is *x*?" isn't the semantic content of that sentence, it is speech act content and the answer given in the form "*x* is *F*" isn't the semantic content of that sentence, it is what is said by it.

In light of this, it might seem that one of Schroeter and Schroeter's questions takes on added urgency: "What is the point of semantic contents?" I have a concessive reply and a dismissive reply to this. The dismissive reply rejects the demand for "points" and insists that truths don't come with points. The concessive reply concedes this talk of points, and provides some of them:

a. We can access semantic contents disquotationally and the knowledge we get that way is substantive.
b. The semantic content of a sentence plays an important role in determining what speakers say by uttering that sentence. The step from semantics to speech act content is messy and unsystematic, but semantics plays an anchoring role. (It's an interesting theoretical project to determine what it is to anchor.)
c. As a corollary of **b**, semantics plays an important constraining role in explaining topic continuity, but, again, it does so in an unsystematic and unpredictable way.[10]

Schroeter and Schroeter will ask: "How can semantic contents play these roles if they are inscrutable?" I have in effect responded to this. First, Schroeter and Schroeter's construal of my talk of inscrutability goes beyond what I had in mind because we have access to semantic content disquotationally (and we don't need "substantive" (nondisquotational) knowledge for it to play these roles.) Second, even if we had a bit of nondisquotational knowledge, it wouldn't make it easier to understand how these "roles" could be performed[11] (so this is really a tangential issue).

Contestation All the Way Down and Nonfactualism about Semantics

According to Schroeter and Schroeter, I "… embrace a version of *nonfactualism* about semantics, metasemantics, and topic continuity. When it comes to meaning, it's contestation all the way down" (15). If so, Schroeter and Schroeter argue, there are no facts about semantics and metasemantics. If there are not semantic facts, then my thesis of inscrutability is wrong: If there are no semantic facts then there are no inscrutable semantic facts.

[10]These are big issues and a full exploration and more elaborate defenses can be found in other work (see *FL* chapter 18 and Cappelen 2019).

[11]See note 8 above.

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

Reply: I'm a factualist about semantics. For example, I think there can be facts about what a name, "*N*," refers to. There are conditions, *C*, such that "*N*" refers to an object *o* just in case it satisfies *C*. *C* will be hard to articulate and we won't be able to do so in a reductive way.[12] That said, we can say informative and true things about *C*, e.g., that it saliently involves causal communicative chains, that it doesn't supervene just on individual speakers' mental states, that it supervenes on use patterns, and that the supervenience relation is extremely complex (and maybe too complex for human beings to grasp). However, and this is relevant in response to Schroeter and Schroeter, that there is such a *C* doesn't mean that it is unchangeable. Two analogies: (1) There are conditions that someone has to satisfy to be polite (relative to a particular contextual setting), and it's a fact that some people are polite and others are not. That's compatible with the standards for politeness changing over time. (2) There are conditions that must be satisfied for someone to be a civilian (in a war) and so it is a fact that some people are civilians and others are combatants. That's compatible with the conditions for being a civilian changing over time (and being negotiated continuously). I think "being the referent of expression *E*" is a bit like that. It's a conventional relation that we, for various reasons, care about and we gradually change it over time[13].

Schroeter and Schroeter's comments touch on a broad range of important issues that I should have been clearer about in *FL* and I want to end this part of the reply with two brief additional comments.

On the Role of the Interpreter's Context

In earlier work, I have argued that what a speaker, *S*, says by uttering a sentence will depend, in part, on the interpreter of *S* (see my 2008). What *S* said can vary depending on the interpreter. I call this view Content Relativism.[14] As I understand Schroeter and Schroeter, they endorse (or at least don't object to) this form of content relativism. What they object to is that topic continuity is interpreter relative in this way (they say, e.g., "The *rationality and successful epistemic coordination* of those engaged in a debate does not vary depending on whether we ourselves are talking to a child or an expert" (14).

This strikes me as an unstable (borderline incoherent) position to occupy. Suppose relative to interpreter *I1*, *A* and *B* have said the same: They are in agreement *that* F*s are* G*s* (both *A* and *B* have said that *F*s are *G*s). Schroeter and Schroeter suggest that we separate topic continuity from samesaying. If so, they suggest it could be true that both *A* and *B* said that *F*s are *G*s, but that they were really talking about different topics. But how can that be when they are both talking about *F*s and *G*s and trying to answer the question, "Are *F*s *G*s?" To deny that there is some very significant sense in which they are then talking about the same topic seems incoherent (assuming you agree, as Schroeter and Schroeter do, that they both said that *F*s are *G*s. (Of course, none of this is to say we can't construct a new meaning for "topic" and, if that's the suggestion, then I say a bit about that option in the reply to Sundell).

On Lack of Control and Flailing About

Finally, Schroeter and Schroeter say, "Cappelen's picture of conceptual engineering is thus a matter of flailing about in the hopes that one will stumble in a better direction" (3). I've noticed that Schroeter and Schroeter are not alone in thinking that *FL* advocates the Flailing About View, but

[12]I.e., we will have to use "reference" or cognate expressions in articulating *C*.

[13]The relevant comparison is not that "*A* refers to *o*" is context sensitive. Even if it is not context sensitive, the relation can change over time. Suppose an expression has a Kaplanian character (variable or not); that character can change over time (maybe as the result of conceptual engineering, i.e., we thought the old Kaplanian character was defective in some way).

[14]Note that this is not relativism about truth: truth on this view can be a monadic property. It is also not nonfactualism: there are interpreter-relative facts.

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

that's not what I had in mind (though given how pervasive the interpretation is, the book no doubt failed to make this clear). Here is why. First, Conceptual Engineering has three parts:

(i) The assessment of representational devices
(ii) Developing proposal for how to ameliorate representational devices
(iii) Efforts to implement the proposals in (ii)

In *FL*, I argue that we lack control over (iii),[15] but that doesn't imply that parts (i) and (ii) are also something we lack control over.

Second, with respect to (iii), and stepping a bit back from Schroeter and Schroeter's reply, the reaction to the book that has surprised me the most is the apparently widespread view that we have a high degree of control over the meaning of English words: that we can develop implementable strategies that, with a high degree of predictability, we can expect to succeed. That view strikes me as prima facie extremely implausible and I'm baffled by the apparent consensus to the contrary. If the meaning of words were easy to control, then English would immediately explode (or implode). There are too many speakers (literally billions, in billions of contexts) with indefinitely many inconsistent preferences, intentions, assessments, goals, plans, and strategies. If English were easy to change, it would collapse. We speakers are fickle, inconsistent, and contentious. Our languages are stable and conservative. The latter is in part what makes the former possible.

Reply to Sarah Sawyer

The Austerity Framework is austere in part because it doesn't rely on a theory of concepts. Surprisingly—and terminologically paradoxically—my account of conceptual engineering has no concepts in it. This, I argue, is good, because concepts are theoretically contentious entities and there's no theory of concepts that I find convincing and philosophically useful. Moreover—and this is an important part of the dialectic in *FL*—even if you disagree with me about that and you fancy a particular account of concepts, you can just add that on later. It would be an add-on, not a replacement. This is the central point at which Sarah Sawyer disagrees with *FL*: Sawyer argues that there's theoretical work at the center of conceptual engineering that cannot be done without an appeal to concepts. Sawyer and I agree that any theory of conceptual engineering should include an account of the limits of revision. That is, any such theory should ask and answer these questions: *How much revision (amelioration or engineering) is too much? When does engineering lead to a change in topic?* In *FL*, I call this "Strawson's challenge" and several chapters of the book are devoted to giving an account of it. Sawyer argues that my account fails and that to fix it I need to introduce concepts. Her argument has two parts: First she provides examples that aim to undermine the account of topic continuity in *FL*. Second, she provides an alternative account that appeals to concepts. Below I address these two parts in turn (even if, as I'm about to argue, the first part of the argument fails, it's important to address the positive proposal: it might still be superior to the Austerity Framework).

Topic Continuity without Overlap in Extensions

According to Sawyer (2020),[16] the theory in *FL* cannot account for cases of conceptual engineering where there's a lack of overlap in extension pre- and postrevision. Suppose there's an expression, *E*,

[15]I should point out that, though I like the expression, "flailing about" does not mean the same as "lack of control." I lack control over how much my students get out of my lectures, but it would be inaccurate to describe myself as "flailing about" in that respect.

[16]In the "Reply to Sarah Sawyer" section, unless otherwise stated, all folios refer to her 2020 article, "The Role of Concepts in *Fixing Language*."

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

that at t has extension S, then a revision happens, and as a result E at t' has S' as extension.[17] According to Sawyer, there can be cases where S and S' have no common elements, but it's still a good example of conceptual engineering that is topic preserving. Sawyer gives several examples of this and I will focus on one of them:

> … the amelioration of race terms can be understood as an attempt to replace a mistaken, biological understanding of race by an explicitly social understanding of race on the grounds that there are no such things as biological races… . In this kind of case, there would be topic continuity in the absence of any overlap in extension since the topic of a term such as "Latino," for example, would be Latinos both preamelioration and postamelioration, even though the conditions that determine the extension of the preameliorative term determine an extension which is empty—empty because there is nothing that satisfies the condition of being a person of biological race. (6)

This case is very helpful to think about and *FL* does not discuss such cases.[18] What follows are some tentative thoughts:

First notice that *FL* doesn't say that the kind of case Sawyer describes is impossible or even unlikely. It is true that the paradigms I use to illustrate topic continuity all involve overlap of extension. I did not, however, argue that this is a necessary condition on samesaying. On the contrary: I emphasised that I *didn't* have as my aim to give necessary and sufficient conditions for samesaying (or topic preservation). As mentioned above in the reply to Sundell, I gave examples of what factors influence topic preservation (genealogy and Railton's checklist can give us an indication of what these are), but throughout I insist on the relation being both open-ended and in flux. I also argue that adding an invariant object as the glue (as what creates unity in diversity) is a mistaken strategy (that's the point of that passage from Williamson I appealed to above).

Of course, none of this explains *why* there is continuity in the particular cases that Sawyer appeals to. In response to that why question, I have one conjecture and two observations. Here is the conjecture (which draws in part on the brief discussion of Railton's checklist in the book [section 10.7]): Suppose, for the sake of argument, we assume that Sawyer's description of the case is correct (i.e., preamelioration, "Latino" had an empty extension while postamelioration it had many people with various characteristics in its extension). An important reason for the continuity is found in the history of this evolution and in particular in the continuity of what most speakers *think* is in extension pre- and postamelioration and in the continuity of what are *taken* to be paradigms. This leads to the first observation: When the conditions described in the conjecture are *not* in place, it becomes hard to defend topic continuity. Imagine, for example, a proposed amelioration of "Latino" that had only blue-eyed Finns in its extension. That doesn't satisfy the conditions in the conjecture and would, in almost any setting, be considered a revision of topic. The second observation is the following: It's not too hard to imagine settings in which Sawyer's case *wouldn't* be considered topic preserving.[19] Imagine a medical setting where the central focus is on how previous talk and thought (in medicine) involving "Latino" involved failed attempts to track a biological category. In those settings, talk that redefined "Latino" as a social (not genetically based) category would be a change of topic; there would be no continuity between the preamelioration inquiry and the issues that arise postamelioration.

Topic Continuity as Conceptual Continuity

According to Sawyer, the only way to account for topic continuity through changes in extension is by introducing a separate representational relation: one that connects a term to a topic. The

[17]Where both S and S' are relative to a fixed point of evaluation, say (w,t^*), E's extension varies pre- and postrevision.

[18]That is, cases in which preamelioration a term had an empty extension. While the example in the text (and others we might think of) might be contentious, no doubt there can be such cases, and they are deserving of our attention.

[19]Remember, according to *FL*, topic preservation can vary between interpreter contexts.

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

Austerity Framework has only one such representation relation (relating terms and extensions and intensions). Sawyer and I are in agreement that there can be topic continuity through changes in extension and intension, but according to Sawyer, this happens because concepts represent topics in ways that "transcend social practices" (9). Sawyer's earlier work (e.g., among others Sawyer 2018) has developed a rich theory of concepts and her reply relies on that theory. Insofar as she has independent arguments for that view, I can't address those here. In what follows, I'll focus on the parts of her theory that are directly relevant to the issues about topic preservation. In that connection, a central thought is that concepts are not determined by social conventions but are "beholden to the nature of the objective properties they represent" (9). These objective properties are features of the world, and so do not vary with things like social conventions (or expert opinion).

My reply has two parts:

- First, I don't need concepts to explain topic continuity. In particular, the cases Sawyer uses to argue against my account and to motivate her alternative (e.g., the "Latino" example above) don't tell against the Austerity Framework (see discussion above). Simplicity considerations thus favor the Austerity Framework.
- Second, Sawyer's view violates the general principle that we should not expect unity in diversity to be grounded in an invariant feature that's present throughout the unified thing (introduced in the reply to Sundell above). In this particular case, the reason is fairly clear: the relation between an expression (e.g., "Latino") and a concept has to be conventional (even if, as Sawyer insists, the relationship between the property and the concept isn't). If so, an expression can shift ever so slightly from one concept to a neighboring one as part of a conceptual engineering process. In such cases, we can have continuity in topic that is not based on continuity of concept. A plausible instance of this is one of Sawyer's examples: the expression "Latino," according to Sawyer, at first picked out an uninstantiated property, and then was revised to denote a (socially constructed) instantiated property. It's hard not to see this as a case where an expression ("Latino") has shifted from one concept (denoting one property) to another concept. However, I should immediately add that I don't have proof of this because we are not told enough about how to individuate concepts associated with expressions and we are not told about the conditions under which an expression can shift between concepts. Maybe Sawyer has an argument either for the view that such shifts are impossible or that they by necessity will coincide with topic change. I'll leave that as an open question.

The Worldly Construal of Conceptual Engineering

In *FL*, I argue that if, for example, "family" has changed its extension and intension between 1819 and now, it is true to say that *what families are* has changed over the last two hundred years. In saying that, we are not just talking about the expression "family"—it is not, I insist, a metalinguistic claim. There are no hidden quotation marks in that sentence. The claim is about families, not about a sequence of letters. In arguing for this I appeal to the view that any utterance expresses a plurality of proposition and the "worldly" reading is just one of the contents that can be expressed by "what families are has changed" (and it is not the semantic content). Sawyer says I have no good argument for that claim:

> I am happy to grant for the sake of the argument that each utterance expresses a large number of propositions only one of which is semantically expressed, but it is unclear how any of the additional propositions expressed by an utterance of "what a belief is has changed" could be true without being metalinguistic. What is needed, after all, is a kind of content that "reflect [s] semantic change" without mentioning the semantic entities (the words) that have undergone the change. (11)

https://doi.org/10.1017/can.2020.8 Published online by Cambridge University Press

Reply: I agree that it is puzzling how we are able to do this, but no more so than many puzzles that come up when we talk about objects whose identity stays the same while they change over time (or across possible worlds). Here is an analogy: Laws change over time and there can be continuity in a law even as it undergoes changes (it is *the same law* that undergoes the changes). What's illegal according L at time t can be different from what's illegal according to L at time t'. In language, we have the flexibility to both talk about L as it is at the time of speaking (e.g., when we say "L makes it illegal to F"), and also to talk about how L has changed over time ("It didn't use to be illegal to F according to L"). There's disagreement about how to account for talk about change more generally, but one view that seems implausible here is that we *have to* give a metalinguistic explanation. We are able to talk about the law itself, not just about the expression "L." This, then, is an illustration of a content that "reflects semantic change" without mentioning the words that have undergone the change. There are many familiar theories about how we talk about change, all of them controversial. One option is this: we need, in effect, a smaller object (e.g, *the law at* t′) and some kind of larger object (e.g, the sequence or "worm" of time slices) and our talk switches between them. The analogous move with respect to talk of meaning change over time is that we have a larger object, *topics* (*family* is an example of a topic) and smaller stages: the semantic values at particular times. The claim in *FL* is that even though topics are not semantic values, we can say things that in complex ways attribute properties to them (e.g., that their parts have changed over time.) That's what we do when we say that, over the last two hundred years, the nature of families has changed.[20]

Herman Cappelen is a professor of philosophy at the University of Oslo and at the University of St. Andrews. He is the author and coauthor of nine books and many papers. His most recent books are *Fixing Language* (2018, Oxford University Press) and *Bad Language* (2019 Oxford University Press, with Josh Dever).

References

Cappelen, Herman. 2008. "The Creative Interpreter: Content Relativism and Assertion." *Noûs* 42 (1): 23–46.

Cappelen, Herman. 2018. *Fixing Language*. Oxford: Oxford University Press.

Cappelen, Herman. 2019. "Conceptual Engineering: The Master Argument." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett. New York: Oxford University Press.

Haslanger, Sally. 2000. Gender and Race: (What) Are They? (What) Do We Want Them to Be? *Noûs* 34 (1): 31–55.

Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Plunkett, David, and Timothy Sundell. 2013. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13 (23): 1–37.

Sawyer, Sarah. 2018. "The Importance of Concepts." *Proceedings of the Aristotelian Society* 118 (2): 127–47.

Sawyer, Sarah. 2020. "The Role of Concepts in *Fixing Language*." *Canadian Journal of Philosophy*.

Schroeter, Laura, and François Schroeter. 2020. "Inscrutability and Its Discontents." *Canadian Journal of Philosophy*.

Sundell, Timothy. 2011a. "Disagreement, Error, and an Alternative to Reference Magnetism." *Australasian Journal of Philosophy* 90: 743–59.

Sundell, Timothy. 2017. "Aesthetic Negotiation." In *Semantics of Aesthetic Judgments*, edited by James Young. Oxford: Oxford University Press.

Sundell, Timothy. 2020. "Changing the Subject." *Canadian Journal of Philosophy*.

Thomasson, Amie. 2019. "A Pragmatic Model for Conceptual Ethics." In *Conceptual Engineering and Conceptual Ethics*, edited by Alexis Burgess, Herman Cappelen, and David Plunkett. New York: Oxford University Press.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Malden, MA: Blackwell.

[20]One might think that we already have larger objects that account for change over time, namely temporal intensions (functions from times to extensions). But that's not what I mean here. If we were to express what I mean in terms of function-theoretic semantics, topics would be something like super- (temporal)intensions: functions from times to intensions.

Cite this article: Cappelen, H. 2020. Conceptual Engineering, Topics, Metasemantics, and Lack of Control. *Canadian Journal of Philosophy* 50: 594–605, doi:10.1017/can.2020.8