SEMANTIC DEFERENCE VERSUS SEMANTIC COORDINATION

Laura Schroeter and François Schroeter

ABSTRACT

It's widely accepted that social facts about an individual's linguistic community can affect both the reference of her words and the concepts (or idiolect meanings) those words express. Theorists sympathetic to the internalist tradition have sought to accommodate these social dependence phenomena without altering their core theoretical commitments by positing deferential reference-fixing criteria. In this paper, we sketch a different explanation of social dependence phenomena, according to which all concepts are individuated in part by causal-historical relations linking token elements of thought.

It's widely accepted that social facts about an individual's linguistic community can affect both the reference of her words and the concepts (or idiolect meanings) those words express. According to Hilary Putnam (1975), experts' criteria for identifying the extension of a natural kind term like "gold" help determine the reference of less expert speakers' use of that term. So, if sameness of meaning for such terms requires sameness of reference, as Putnam believes, then the individuation of non-experts' idiolect meanings will depend on contingent facts about their social environment. According to Tyler Burge (1979), facts about an individual's linguistic community can affect the reference of the *thoughts* expressed by her non-indexical expressions. So, if sameness of non-indexical concepts requires sameness of reference, as Burge believes, then the individuation of most concepts will depend in part on contingent facts about an individual thinker's social environment.

Putnam and Burge took these social dependence claims to constitute a radical departure from traditional accounts of the determination of reference and the individuation of representational state types. In response, theorists sympathetic to the internalist tradition have argued that they can accommodate the type of social dependence phenomena highlighted by Putnam and Burge without altering their core theoretical commitments. All that Putnam and Burge have established is that some concepts are *deferential*: other speakers' ways of using words may figure in an individual's own criteria for identifying the reference. On this deferential account of reference determination, both the application conditions of a subject's words and the meanings and concepts those words express are determined in a straightforward way by her internal states.

In this paper, we sketch an alternative model of how social dependence relations affect meanings, concepts, and reference. On the binding model of concept identity, social dependence relations are part of the basic infrastructure for concept identity. We start in section 1 by exploring how an internalist

might accommodate social dependence through deferential reference-fixing criteria. In section 2, we propose two core constraints on an adequate theory of concept identity. We then contrast the traditional, broadly Fregean model of concepts with the binding model (section 3), and explain the different consequences the models have for who shares concepts (section 4). In closing, we offer two reasons to think the binding model is a promising alternative to the traditional internalist approach to representation (section 5).

ɪ. THE ACCOMMODATION STRATEGY: SOCIAL DEFERENCE

We take *internalism* to be a metasemantic thesis about which factors contribute to determining the semantic values of a subject's words and concepts. Our focus here is on the concepts expressed by context-invariant referential expressions (such as "elm," "arthritis," "Gödel," "water," "sofa"), which have been the target of externalist arguments.[1]

> Internalism: For any token element of thought [*x*], there are associated internal states (i.e., structural, dispositional, or phenomenal states individuated independently of the subject's actual external environment) that (i) constitute the subject's core understanding of its reference, and (ii) provide sufficient constraints to fix its reference in all possible worlds.

In other words, subjects rely on a *reference-fixing criterion* that would put them in a position to identify which things fall into the extension of the concept in any possible world on the basis of ideal rational reflection: if presented with a full description of a possible world, the subject would in principle be able to identify precisely which things fall into the extension.[2]

The core understanding that constitutes the subject's reference-fixing criterion can be realized in different psychological states. According to simple descriptivism, for instance, a reference-fixing criterion involves the ability to define the essential nature of the object, kind, or property picked out. More sophisticated forms of internalism allow for implicit reference-fixing criteria, which may be embodied in recognitional or inferential dispositions, beliefs about paradigm exemplars, prototypes composed of weighted sets of properties, internal mental models, and so on. Just which states embody the reference-fixing criterion is determined by which dispositions ultimately guide the subject when identifying the reference under conditions of ideal rational reflection.

Externalist thought experiments provide plausible counterexamples to simple descriptivism. According to our best reflective judgments, a subject can be ignorant or mistaken about the nature of the objects, kinds, or properties she picks out in thought. Moreover, these mistakes can often be corrected only by learning empirical facts about one's environment. Despite Putnam's inability to distinguish elms from beeches on the basis of their observable physical properties, for instance, we take his use of "elm" to refer determinately to elms. Part of the explanation for how this determinate reference is secured, Putnam suggests, is that ordinary speakers like himself are disposed to defer to "experts" in their linguistic community about what counts as an elm (1975). And in Burge's thought experiment, we take Bert's thoughts to determinately pick out the very same property as the rest of us—arthritis—despite Bert's mistaken belief that arthritis can occur in the thigh. Part of what justifies this attribution, according to Burge, is that Bert would take himself to stand corrected if his doctor (or a dictionary) informed him that arthritis is a disease of the joints (1979). If we take these judgments about reference at face value, it seems that facts about the subject's social environment can be crucial to determining the reference of her words and thoughts. In that case, the subject's current reference-fixing criteria won't always suffice to identify the reference of her words and

thoughts independently of contingent facts about her actual social environment.

There is a standard reply to such externalist arguments. The very fact that subjects take themselves to stand corrected by expert opinion reveals that deference to experts (or some other social factor) is part of their ultimate criterion for identifying the reference. If presented with a full description of a possible world that includes information about the social environment, the subject would be able to identify which things fall into the extension of the target words and concepts after ideal reflection—independently of knowing any empirical facts about her actual environment.[3] For instance, Putnam's criterion for being an elm may be: whatever satisfies the criteria the relevant experts associate with "elm." This deferential criterion fixes the reference of Putnam's term *indirectly*, as a function of contingent facts about his actual social environment: different properties will be picked out, depending on the opinions of experts in one's social community. The internalist can then use the referential judgments elicited by social externalist thought experiments to specify a subject's deferential reference-fixing criterion.[4]

One might worry that any specific deferential analysis will be subject to counterexamples. The good news for the internalist is that one can appeal to the 2-D semantic framework to specify a subject's internal reference-fixing criteria (including her ultimate dispositions to take facts about her linguistic community into account) without needing to formulate a strict, counterexample-free definition in natural language.[5] According to David Chalmers and Frank Jackson, the reference of a token use of a word "*x*" (and of the concept thereby expressed) is determined by the individual's dispositions to judge which things are *x*, given full empirical information about base-level facts about a possible world considered as actual and ideal cognitive powers.[6] Chalmers and Jackson then use a 2-D framework to

define an internally determined semantic value that marks these dispositions: the *epistemic intension* of a subject's term "*x*" is a function from every epistemically possible empirical scenario (roughly a centered possible world) to the subject's ideal verdicts about what *x* is relative to that scenario considered as actual. Given that the epistemic intension is determined by the subject's purely internal states together with ideal armchair reasoning about hypothetical scenarios, the epistemic intension can be used to demarcate a pattern of understanding that suffices to fix the reference *come what may*. On this approach, a subject's reference-fixing criterion is simply whichever pattern of actual understanding would ideally generate a specific epistemic intension.

The key novelty introduced by social dependence phenomena in the 2-D framework is that the subject's epistemic intensions take into account not just facts about her external physical environment, but also *meta-cognitive facts* about others' understanding and use of "*x*." The subject's ideal dispositions to react to meta-cognitive information will reflect her ideal views about who counts as members of her linguistic community, about which events count as uses of the same term "*x*," about how to take others' opinions about the applicability of "*x*" into account, and about when to conclude that others' opinions diverge too much to be relevant to fixing the reference of her own use of "*x*."

The availability of this 2-D strategy suggests that externalist counterexamples (social or otherwise) needn't pose any deep challenge to metasemantic internalism. Insofar as the alleged counterexamples accord with the subject's own ideal reflective verdicts, there will be internal dispositions that generate those verdicts. These dispositions can then be captured by an epistemic intension. And insofar as alleged counterexamples flout the subject's own ideal dispositions, the internalist can simply deny that they have any intuitive force.

2. CONCEPTS AND APPARENT *DE JURE* SAMENESS

To mount a plausible challenge to the traditional approach to representation, a semantic externalist should not focus on the internalist metasemantics considered in isolation. Instead, she should step back and ask how the internalist's accommodation strategy *fits within a broader theory of representational state types*. The key question is how an internalist metasemantics, which invokes social deference, contributes to a plausible account of concepts (and idiolect meanings).

A theory of concepts must answer two questions, one about what it takes to instantiate the same representational state type, and one about the determination of semantic values:[7]

1. <u>Concept identity:</u> What makes it the case that two token words or elements of thought express the same concept?
2. <u>Metasemantics:</u> What makes it the case that a token word or element of thought has the semantic value it does?

Focusing on representational state types reflects core aspects of the philosophical debate over metasemantic internalism. Early proponents of social dependence relations like Kripke, Putnam, Devitt, and Burge had a guiding interest in individuating shared representational state types. In particular, the most influential proponent of social dependence relations, Tyler Burge, has repeatedly emphasized that his primary concern is to show how external factors are crucial to determining whether two individuals share the same fine-grained representational states—concepts or whole thought contents.[8] Likewise, metasemantic internalists like Lewis, Jackson, Chalmers, and Segal highlight the role an internalist account of reference-determination plays in individuating idiolect meanings and concepts.[9]

In this section, we'll clarify two constraints on an adequate theory of concept identity.

This step is important because it will put us in a position to articulate and motivate an alternative to the traditional theory of concepts in which metasemantic internalism finds its home. To make room for this alternative approach, we'll need to isolate the precise explanandum in a theory of concept identity. On the alternative binding model of concepts that we'll sketch, social relations play an essential role in concept individuation—one that's very different from the reference-fixing role assigned to social dependence relations by the internalist accommodationist strategy canvassed in section 1.

A first constraint on an adequate account of concept identity is that, for non-indexical concepts like [elm], *sameness of concept entails sameness of reference*. Any two individuals who share the concept [elm] must thereby represent the very same property—no matter how different their background beliefs about the topic, and no matter how different their external social or physical circumstances. So the theory of concept identity and the metasemantic theory, together, must ensure that any two tokens of the concept [elm] must be assigned the very same reference, irrespective of the psychological, social, or physical contexts in which those two tokens occur.

A second constraint concerns the cognitive significance marked by concepts. It's generally agreed that concepts mark modes of presentation or ways of thinking of reference. More specifically, we suggest, the core explanatory role of concepts is to capture our most basic ways of keeping track of a topic in thought. Ordinary human learning and reasoning would be impossible without some basic ways of organizing our cognitive states by topic. Our ability to learn about Hesperus, for example, to amass a body of attitudes and dispositions pertaining to that topic, to form new beliefs about it, to retrieve relevant information *via* memory, and to stay on topic in reasoning—all these basic cognitive abilities depend on a stable organizational structure.

We have argued elsewhere that these basic ways of organizing our cognitive states by topic depend on subpersonal mechanisms that give rise to a specific epistemic signature at the conscious level.[10] When a subject consciously entertains a train of thought, certain elements *seem guaranteed* to pertain to the very same topic, simply in virtue of the way they are presented in thought. We call this the appearance of *de jure* sameness. So, concepts must be individuated in such a way as to *reflect the subjective appearance of guaranteed sameness of topic*.

To get a clearer view of the appearance of *de jure* sameness, consider an example. Throughout a conscious episode of reasoning about Hesperus, a subject's thoughts present themselves to her conscious attention as obviously and incontrovertibly about the very same thing, Hesperus. We'd like to emphasize three features of this apparent *de jure* sameness:

i. Obviousness: The subject is *immediately* disposed to treat these thoughts as pertaining to the same topic in her reasoning. There's no need for her to engage in reflection or investigation to establish sameness of topic. Indeed, in entertaining these thoughts, the question of sameness simply doesn't arise: from the object-level perspective, it's not an open question whether Hesperus is Hesperus.

ii. Rational incontrovertibility: From the object-level perspective, it seems *logically incoherent* to deny the identity of topic: there is no conceivable circumstance in which Hesperus could fail to be Hesperus. (A logically coherent doubt can, however, be raised if one adopts a meta-level perspective on one's own thoughts: one can coherently deny that certain thoughts co-refer.)

iii. Epistemic basicness: This subjective appearance of obvious and incontrovertible sameness is *not based on or reinforced by* independent evidence of sameness: no object-level evidence about Hesperus could help justify the apparently trivial identity claim that Hesperus is Hesperus.[11]

It's important to note that this characterization of apparent *de jure* sameness does not entail that the appearance is veridical. Indeed, the appearance of *de jure* sameness is only characteristic of the subject's epistemic perspective in object-level thinking about the world—but it disappears as soon as one adopts a meta-level perspective on those thoughts.

Not all thoughts that the subject takes to pertain to the same topic exhibit apparent *de jure* sameness. Unlike the question of whether Hesperus is Hesperus, the question of whether Hesperus really is Phosphorus is open: one could rationally doubt the identity and one could cite object-level empirical facts that would lend support to one's conviction that there's just one topic in question. We call the phenomenology associated with non-trivial identities the appearance of *de facto* sameness of topic.

The distinction between apparent *de jure* and *de facto* sameness, we suggest, captures the *core explanandum* of Fregean cognitive significance: the subjective epistemic difference between trivial and non-trivial identities in thought. Unlike standard ways of understanding Fregean cognitive significance, however, the notion of apparent *de jure* sameness does not build in any assumptions about what generates or vindicates these subjective appearances. In particular, it does not assume that the appearance of *de jure* sameness is explained by a match in the reference-fixing criteria associated with token words or elements of thought.

The appearance of *de jure* sameness is crucial to determining direct logical relations among thought contents. The premise [Hesperus appears in the evening], for instance, does not logically entail the conclusion [Phosphorus appears in the evening]: minimal logical coherence does not require the subject to accept the conclusion whenever she accepts the premise, even though the truth of the premise metaphysically guarantees the

truth of the conclusion. Similarly, relations of apparent *de jure* sameness are necessary for direct logical contradictions among thought contents. The fact that it's logically contradictory for the subject to accept that Hesperus appears in the evening and to accept that Hesperus does not appear in the evening depends on the apparent *de jure* sameness of topic: from the subject's perspective, it seems obvious and incontrovertible that both thoughts pick out the very same thing (Hesperus) and attribute the very same property (appearing in the evening). Of course, genuine logical relations among thoughts also require *objective* sameness of topic, not just the *appearance* of sameness. But when appearances of *de jure* sameness of topic are veridical—that is, when they're vindicated by the correct metasemantic theory—then the resulting *de jure* sameness relations fix the *logical form* of thought contents. A primary role of a theory of concept identity is to mark these relations of *de jure* sameness among elements of thought.

The role of concepts in demarcating the logical form of thoughts underwrites the core theoretical roles they play in rationalizing explanations and in normative assessments of the subject's rationality. The subjective appearance of *de jure* sameness by itself helps explain why subjects are disposed to combine certain standing attitudes in reasoning. If a subject consciously entertains thoughts of the form $[Pa \rightarrow Gb]$ and $[Pa]$, then given the appearance of *de jure* sameness among the constituents, we'd expect her to draw the conclusion $[Gb]$. These subjective appearances also determine norms of minimal rationality: if the subject were to consciously accept thoughts of the form $[Gb]$ and $[\sim Gb]$, we'd criticize her for flouting standards of minimal logical coherence. Finally, the veridicality of apparent *de jure* sameness relations is crucial to determining the subject's ultimate rational goals and corresponding standards for normative assessment. If the appearance of *de jure*

sameness linking the elements of the subject's modus ponens inference were not veridical, then we would not assess her reasoning as logically valid: despite the subjective appearance of validity, the reasoning would be objectively invalid. The objective perspective on logical relations is important because it captures the subject's own commitments to getting facts about the world right, and the assessor's interest in relating the subject's thoughts to the world.

We have highlighted the importance of apparent *de jure* sameness as the epistemic signature of our most basic ways of keeping track of sameness of topic in thought. A theory of concepts should individuate representational state types in such a way as to reflect these subjective appearances. At the same time, to capture genuine logical relations among thoughts, a theory of concept identity together with the metasemantics must explain how sameness of non-indexical concepts guarantees objective sameness of reference. This dual role of referential concepts, in marking both the appearance and reality of *de jure* co-reference, sets up important challenges, especially in explaining concept identity over time and between subjects. In the next section, we contrast two models of concept identity, the matching model and the binding model, which represent two strategies for answering the metasemantic and identity questions in a theory of concepts.

3. TWO MODELS OF CONCEPT IDENTITY

The traditional broadly Fregean approach to concepts combines a *matching model* of concept identity with an internalist metasemantics. We will start with a simple version of the model:

1. Matching model$_1$: Two token elements of thought $[x]$ and $[y]$ express the same non-indexical referential concept iff each token is associated with exactly the same reference-fixing criterion C_n.

2. Internalism: For any token element of thought [*x*], there are associated internal states *F* that (i) constitute the subject's core understanding of its reference, and (ii) provide sufficient constraints to fix its reference in all possible worlds.

According to this matching model, a single theoretical posit—a reference-fixing criterion associated with a token element of thought at a given time—fulfills both theoretical roles in an account of referential concepts. Each token element of thought is independently associated with a reference-fixing criterion, C_n, that suffices to determine the essential nature of the object, kind, or property represented. And concept identity is determined by a precise match in reference-fixing criteria: two elements of thought express the same concept just in case each is associated with the very same reference-fixing criterion C_n.

The matching model of concepts suggests a straightforward way of satisfying our two constraints on a theory of concept identity. First, concepts must be individuated in such a way as to reflect the subjective appearance of *de jure* sameness. On the matching model of concept identity, sameness of concept is determined by internal states that constitute the subject's implicit understanding of the reference. If two tokens of [Hesperus] are each associated with precisely the same implicit reference-fixing criteria, this fact might explain why it seems obvious and incontrovertible that there's just one thing in question when one entertains those tokens in one's object-level thinking. Second, for non-indexical referential concepts like [Hesperus], sameness of concept entails sameness of reference. If the criteria associated with a token element of thought suffice to determine its reference in any possible world, then the matching model of concept identity can explain why any two tokens of the same concept will be logically guaranteed to co-refer: not only will two tokens associated with the same reference-fixing criteria *seem* co-referential,

according to internalist metasemantics, they are guaranteed to co-refer irrespective of the empirical context in which they occur.

However, the indirect reference-fixing strategy that internalists appeal to in order to accommodate externalist counterexamples introduces a complication to the matching model of concept identity. An indirect reference-fixing criterion fixes the reference relative to contingent facts about the subject's historical, social, or physical context. So there's no guarantee that two tokens associated with the same indirect criterion must co-refer—for the tokens may occur in different contexts. While the matching model of concept identity may explain why two tokens of non-indexical referential concepts like [Hesperus], [water], or [arthritis] *seem* obviously and incontrovertibly co-referential when one consciously entertains them in thought, it cannot explain why they are *in fact* logically guaranteed to co-refer. The worry is that indirect reference-fixing criteria cannot explain the direct logical relations among thoughts that concepts are supposed to mark.

This isn't a decisive objection to the matching model of concept identity. For the matching model can be modified to accommodate this empirical fallibility while still explaining why the appearance of *de jure* sameness is a reliable guide to the truth. The first step is simply to add an extra clause in the account of concept identity:

> Matching model$_2$: Two token elements of thought [*x*] and [*y*] express the same non-indexical referential concept iff (i) each token is associated with exactly the same indirect reference-fixing criterion C_n, *and* (ii) both tokens occur in contexts that are alike with respect to the relevant empirical reference-fixing features.

So, according to the matching model$_2$, Bert and his counterfactual twin may have different concepts, [arthritis] and [tharthritis], despite the fact that they rely on precisely the same deferential reference-fixing criterion. In

such cases, sameness of concept depends on sameness of social context—just as Burge argues. However, in contrast to Burge's own brand of anti-individualism about concept identity, the matching model$_2$ preserves the core of the traditional matching model of concepts. On the matching model$_2$, the logical guarantee of co-reference is explained in terms of a precise match in competent subjects' reference-fixing criteria, which guarantees two tokens will co-refer relative to the same empirical context—no matter what that context is like. But unlike the traditional matching model, this guarantee can be defeated if the empirical contexts of use diverge, like those of Bert and Twin Bert.[12]

This concludes our rough sketch of how a matching model, broadly construed, seeks to explain both apparent and objective *de jure* sameness relations among thoughts. But there are many details that an adequate model of concept identity must fill in. In particular, we'd like to flag the following outstanding issues for matching models:

1. Apparent *de jure* sameness: In our sketch of the matching model, we focused on explaining the rational *incontrovertibility* that's characteristic of logical relations among thoughts. But, as we saw in section 2, apparent *de jure* sameness involves two further epistemic features that will be harder for a matching model to explain: *obviousness* and *basicness*. And this sets up a challenge for the matching model: the more the model seeks to accommodate externalist intuitions about the ways the subject's current understanding of the reference is fallible, the less subjectively obvious the ultimate reference-fixing criteria will become, and the harder it will be to capture obviousness and basicness. The worry, then, is that the structure of the matching model makes it difficult to fully explain the appearance of *de jure* sameness (Schroeter 2003; Schroeter and Schroeter 2009).

2. Demarcating competence conditions: What are the principles for determining precisely which aspects of associated understanding get included in the reference-fixing criterion C_n for a token element of thought? Can the account avoid an implausible holism in the reference-fixing criteria, which would entail that concepts are not normally stable over time and between subjects? Do we really need to share precisely the same reference-fixing criteria in order to ensure direct logical relations over time and between subjects? Many self-described externalists have raised worries along these lines for matching models of meanings and concepts (Stalnaker 1993; Yablo 2000; Schroeter 2003, 2004, 2006).

3. Deference relations: How exactly do social dependence relations work: Which uses does the subject treat as authoritative, and in what ways? Can social dependence dispositions always be interpreted in a *non-circular* way—so that *A*'s dispositions to defer to an individual or group *B*, and *B*'s dispositions do not take *A*'s dispositions into account? The deference model seems to require that social deference criteria depend for their reference on other's subject's non-social criteria. But, arguably, even the most expert speakers are disposed to take communal use into account in determining the precise reference of their words and thoughts (Burge 1986, 1989; Schroeter 2014; Schroeter and Schroeter 2015).

4. *Apriority* and fallibility: According to a matching model, the subject's ultimate reference-fixing criteria have the same *a priori* epistemic status as a stipulative definition. But how independently plausible is this alleged *a priori* status? (Kripke 1980; Speaks 2010). Can failsafe reference-fixing criteria be reconciled with the idea that we represent objectively important, mind-independent features of environment? (Millikan 1984, 1993).

We raise these points here not as decisive objections to a matching model, but just as an indication that this standard account of concept identity faces significant challenges.

Let's turn now to the binding model of social dependence relations. What's distinctive

of a binding model is that it treats real-world causal-historical relations as at least partly constitutive of concept identity. Here's a simple version:

> Binding Model$_1$: Two token elements of thought [x] and [y] express the very same non-indexical referential concept iff those elements stand in causal-historical relation R to each other.

Sameness of concept, on this account, is determined by a causal-historical relation linking token elements of thought, rather than by a match in the properties independently associated with particular tokens. The intuitive idea is that R binds token elements of thought into a continuous *representational tradition* in which the R-linked words and thoughts are treated as pertaining *de jure* to the same topic. So expressing the same concept is an irreducibly relational fact, like belonging to the same family or the same biological species.

To play the role of demarcating a shared representational tradition, R must include both intrapersonal relations, R_i, linking elements of an individual subject's thoughts at a time and over time, and interpersonal relations, R_s, linking thoughts of different individuals within a linguistic community. For instance, when a subject engages in a chain of conscious reasoning, subpersonal cognitive mechanisms ensure that certain occurrent thoughts are all presented to her conscious attention as pertaining *de jure* to the same topic. Moreover, most of those occurrent thoughts are causally linked to stable dispositional attitudes that persist over time, forming an uninterrupted causal-historical link between her occurrent thoughts now to the past attitudes that originally gave rise to them. The subject might form a stable belief, for instance, that she'd express as "Pavarotti's favorite drink is water"—and this persisting belief will strike the subject as pertaining *de jure* to the same object and the same property throughout its existence. Similarly, when a subject converses with others, her "water" thoughts are causally linked to the thoughts her interlocutor expresses with "water," and chains of such links extend out from her interlocutors to other English speakers, past and present, to form a continuous shared representational tradition.

How can a binding model satisfy our two constraints on concept identity? First, concepts must reflect the subjective appearance of *de jure* sameness. The binding model claims that apparent *de jure* sameness is an irreducibly relational matter: apparent sameness does not supervene on the properties of the bound elements considered independently. Whereas a matching model seeks to explain apparent *de jure* sameness in terms of matching reference-fixing criteria, a binding model treats apparent *de jure* sameness as an epistemically basic fact generated by subpersonal mechanisms linking the relevant token states.

Second, a binding model requires a very different explanation of how sameness of concept guarantees co-reference, from that of a matching model. Given that concept identity is determined by relation R, it's natural to take the default unit for semantic interpretation to be the representational tradition as a whole, rather than token thought episodes considered independently.

> Metasemantic coordination: The default units for semantic interpretation are the temporally and socially extended representational traditions demarcated by R. So, a token element of thought [x] normally inherits its semantic value from the representational tradition to which it is bound by R.

Metasemantic coordination is a constraint on acceptable principles for assigning semantic values on a binding model; it must be supplemented by substantive metasemantic principles. But the coordination constraint is sufficient to distinguish the metasemantic approach characteristic of binding models from that of matching models. On a binding

model, the reference of a token element of thought [*x*] is determined on the basis of a set of properties *P* of the whole representational tradition demarcated by *R*: for example, patterns of attitudes and cognitive dispositions associated with the concept, history of how uses of the concept are causally triggered by features of the environment, causal feedback relations entrenching the use of the concept over time, shared practical and theoretical interests subserved by the tradition, principles of rationalizing interpretation.

If Metasemantic coordination is correct, it's no accident that two speakers of English at a given time refer to the same thing when they think about "water" or "arthritis" or "Gödel": co-reference is ensured by the fact that each token is appropriately causally linked to a common representational tradition. Different speakers might have slightly different commitments about what kind of stuff water is, which samples count as instances of water, the best methods for discovering its nature, or which experts are the most trustworthy, and so on. And these differences might ground distinct reference-fixing criteria for the individuals considered in isolation, yielding different verdicts about the reference relative to some possible social or physical environments. Still, if the default unit for assigning reference is the representational tradition as a whole, rather than a token element of thought at a time, a single univocal reference will normally be assigned to the tradition despite these variations in individuals' understanding.

However, the simple binding model must be modified if it is to accommodate commitments about reference elicited by externalist thought experiments. The problem is that, according to our best reflective standards, the appearance of *de jure* sameness does not guarantee co-reference. Gareth Evans's Madagascar case (1973) and Tyler Burge's slow switching case (1988), for instance, are examples of stable representational traditions

that preserve the appearance of *de jure* sameness over time but which seem to shift their semantic values as a consequence of a persistent mistaken identity: eventually some token thoughts expressed by "Madagascar" and "water" in the tradition come to have a new reference in virtue of new patterns of use by a whole community or by a particular individual. So, the default presumption of co-reference within a single representational tradition can be defeated.

To accommodate such cases, a plausible binding model of concept identity must be hedged, just as the sophisticated matching model was. Here's one hedging strategy:

Binding model$_2$: Two token elements of thought express the same non-indexical referential concept iff (i) those elements stand in a specific causal-historical relation, *R*, to each other, and (ii) there are no undefeated defeaters to the default presumption of co-reference.

On this hedged version of the model, concept identity is beholden to *both* causal-historical facts about representational traditions and metasemantic facts about the best interpretation of those traditions.[13]

We have provided a schematic characterization of the binding model of concepts. The two core commitments of such a model are: (i) that sameness of concept (or meaning) depends on a causal-historical relation *R*, and (ii) that *R* demarcates default units for semantic interpretation. A number of very different theories arguably satisfy these constraints: for example, causal-historical theories of names and natural kind terms (Kripke 1980; Devitt 1981), teleosemantic theories of mental content (Millikan 1984, 2000; Dretske 1988), mental file models of semantic coordination (Perry 1980, 2001; Heim 1982; Cumming 2013), relational semantic rules (Fine 2007; Pinillos 2011), relational accounts of word individuation (Kaplan 1990; Fiengo and May 2006), and our connectedness model of concept identity (Schroeter 2008, 2012;

Schroeter and Schroeter 2009, 2014b). An adequate binding model must flesh out this core structure. Key challenges include:

1. <u>Apparent *de jure* sameness:</u> What exactly is relation R? How is R related to the subjective appearance of obvious and rationally indubitable sameness characteristic of concepts? What exactly are the limits of the representational traditions demarcated by R? One might worry that an account that appeals to socially and temporally extended chains of apparent *de jure* sameness will lose the requisite connection to immediate subjective appearances.

2. <u>Metasemantics:</u> What are the principles for assigning a semantic value to representational traditions? When (if ever) is the default presumption of co-reference defeated? Can such defeating conditions be specified without assuming that token elements of thought earn their semantic values independently of a representational tradition (and thus violating Metasemantic coordination)?

3. <u>Identity:</u> Can a binding model define concept identity as an equivalence relation—that is, reflexive, symmetrical, and transitive? Some theorists have argued that relational accounts will violate transitivity (Fine 2007; Pinillos 2011).[14]

4. DEFERENCE VERSUS COORDINATION

Social dependence relations play a very different structural role in the binding model of concepts than in the matching model. In the binding model, the diachronic and social coordination relations R_i and R_s that define shared representational traditions help constitute concept identity in general. The matching model might include a relation like R_s as part of an individual's deferential reference-fixing criterion C_n for certain token elements of thought. In that case, a disposition to rely on R_s will be a necessary condition for sharing the same deferential concept. But a binding model treats actually standing in relations R_i and R_s as necessary conditions for sameness

of any concept over time or between subjects: these external coordination relations are part of the basic infrastructure for concept identity, not an optional feature characteristic of particular concepts.

This structural difference between the two models generates two very different pictures of who shares socially dependent concepts.

i. According to the binding model of concept identity, individuals who aren't causally connected in the relevant way by R cannot share the same concepts. So, Davidson's swampman, for instance, won't share any of the same concepts as the rest of us; and we won't share any concepts with causally unconnected linguistic communities like those on Putnam's Twin Earth.

ii. According to the matching model of concept identity, individuals who don't share precisely the same reference-fixing criteria C_n cannot share the same concept. So, Putnam and the tree experts to whom he defers don't express the same concept with "elm"; and the concept Putnam himself expresses with "elm" may change, perhaps many times, as he learns more about the nature of elms, about the best methods for identifying them, and about which experts to trust and how much to trust them.

In short, a binding model makes concepts unstable across different representational traditions, whereas a matching model makes concepts unstable across differences in reference-fixing criteria within a single representational tradition.

For many theorists, (i) will seem a problematic consequence of the binding model—for it seems to conflict with commonsense intuitions about sameness of concepts. However, commonsense verdicts about when it's appropriate to *say* people share the same "meaning" or "concept" are sensitive to many different pragmatic factors, and they reflect different theoretically important features of subjects' representational states. Any systematic theory of concept identity will violate some of our commonsense ways of speaking. As (ii)

illustrates, matching models simply flout a different set of commonsense verdicts about who shares the same "concepts." Whether to accept the binding model will come down to the theoretical work it does, not how it conforms to commonsense ways of speaking (Schroeter and Schroeter 2014a).

Let's turn now to (ii). It's well known that the matching model of concept identity faces challenges in accounting for conceptual stability through theory change. But it's important to see that appealing to deferential reference-fixing criteria does not address this problem—if anything, it makes concepts more unstable. Consider a simple deferential model on which the reference of Putnam's "elm" thoughts is determined by his treating a certain type of people, "the tree experts," as infallible gurus about which things count as elms. To share the same deferential concept with Putnam, you must be disposed to treat precisely the same descriptively characterized set of people as gurus. That is, if we were to slot you and Putnam into different possible social scenarios, the two of you would always arrive at precisely the same verdicts about whose word to trust. Even if the two of you happen to defer to the very same set of people in the *actual* world, and so your words have the same extension in the actual world, the fact that your verdicts would diverge with respect to some *non-actual* social context suffices to establish that your reference-fixing criteria diverge. According to a matching model of concept identity, it follows that you do not share the same concept.

This conceptual instability is not an artifact of the particular deferential model we choose to consider. Any non-circular deferential reference-fixing criterion, which makes the reference of one person's thoughts depend on that of others, will have the same consequence.[15] Introducing deferential reference-fixing criteria is incompatible with sameness of concept among all members of a linguistic community.

It's easy to be confused about this point because deferential concepts combine two distinct mechanisms for ensuring co-reference: (a) concept identity, and (b) real-world causal-historical deference relations linking distinct concepts. On the matching model, a match in reference-fixing criteria (together with a match in context) is the general metasemantic mechanism that explains *de jure* co-reference, both for deferential and non-deferential concepts. Given the same physical and social context, it's logically impossible that two individuals with the same reference-fixing criteria could pick out distinct things, since their reference will co-vary no matter what their shared context is like. On this model, direct logical relations among thoughts are secured by matching reference-fixing criteria—these are what determine the logical form of thought contents.

But in the case of deferential concepts there is a further mechanism that can establish co-reference between thinkers. Deferential reference-fixing criteria make the reference of the concept expressed by one person's use of a word depend on the reference of the concepts expressed by others' uses of the word. In effect, a deferential reference-fixing criterion gives instructions for identifying a different, non-deferential reference-fixing criterion and treating it as authoritative. On this view, just which non-deferential criterion is actually authoritative will depend on contingent facts about the subject's social environment. For instance, the reference of Bert's "arthritis" concept will depend on which concepts happen to be deployed by expert speakers within Bert's actual linguistic community. This one-way dependence relation can establish *de facto* co-reference between Bert's deferential concept and the non-deferential concept expressed by those to whom he defers. But there is no *de jure* guarantee of co-reference between a deferential concept and the non-deferential concepts to which the subject happens to defer. It's a

purely contingent matter that Bert's deference dispositions happened to treat as authoritative the non-deferential concept [arthritis] (i.e., rheumatoid ailment of the joints) rather than the concept [tharthritis] (i.e., any rheumatoid ailment). In sum, whereas the matching model allows social dependence relations to play a role in ensuring *de facto* co-reference among members of a linguistic community, they do not ground the *de jure* co-reference relations that are distinctive of concept identity.

5. TWO ADVANTAGES OF THE BINDING MODEL

We have contrasted two ways of incorporating social dependence relations in a theory of concepts—the matching model and the binding model. Although we favor the second model, we have not argued for it here: our aim has been to clarify the core structural differences between the two models. In closing, we would like to quickly highlight what we take to be two important advantages of the binding model—one psychological, one epistemological.

Prima facie, the binding model fits better with both the first-person phenomenology of *de jure* sameness and the cognitive mechanisms that underlie it. The appearance of *de jure* sameness is not confined to co-conscious thoughts within a given individual: it extends to past thoughts and to thoughts linguistically expressed by others. When Bert remembers glimpsing Cate Blanchett at the airport, Bert's occurrent memory seems to present the content of his past thoughts as *de jure* pertaining to the same topic as his current beliefs about the star. And when he hears others using the term "arthritis," he understands their claims as pertaining *de jure* to the same topic as he himself associates with that term. In such cases, the appearance of *de jure* sameness does not seem to depend on a precise match in reference-fixing criteria. First, a subject's ultimate reference-fixing criteria can depend on highly abstract, non-obvious dispositional states like his ultimate dispositions to defer to certain sorts of people under certain conditions, but not others. It's hard to see how the appearance of *de jure* sameness could be generated by detecting a perfect match in such dispositions. Second, it seems possible to coherently reject virtually any particular substantive claim about "arthritis" without threatening the appearance of *de jure* sameness.[16] When Bert's doctor says "Arthritis is a disease of the joints," or when his guru says "Doctors are not experts on arthritis," Bert will automatically hear both claims as pertaining *de jure* to the same topic as his own "arthritis" thoughts. It seems psychologically implausible that this appearance counterfactually depends on a match in criteria that are sufficiently detailed to guarantee co-reference in all possible situations. It's more economical to suppose that subpersonal binding mechanisms generate these appearances directly, without needing to monitor and match the precise associated dispositions.[17]

The binding model also fits better with our normative commitments as rational epistemic agents. Concept identity should not be tied too closely to topic-specific assumptions (criteria of application, epistemic methods for self-correction, and so on) if we want to allow for the possibility of conceptual stability through open-ended inquiry, debate, and disagreement. The matching model individuates concepts in terms of a specific aspect of substantive understanding that's sufficiently rich to fix the reference (relative to a context). As a consequence, the model is structurally committed to strict limits on what can be coherently rejected. On a matching model, successful rational inquiry can easily lead to a change in concept when the subject abandons aspects of her prior understanding that she discovers to be false (Schroeter 2006). So, accepting the upshot of successful rational inquiry can break the chain of *de jure* sameness relations that

ground direct logical relations among thought contents. This verdict doesn't just violate the immediate subjective appearance of *de jure* sameness. From an epistemological perspective, it imposes implausible epistemic norms: in order to maintain direct logical relations with one's past beliefs about a topic (many of which persist as standing attitudes about that topic), one must not abandon certain assumptions that one has discovered to be false. On the binding model, in contrast, sameness of concept does not depend on accepting any particular set of substantive commitments about the reference. Because direct binding relations demarcate stable default units for semantic interpretation, there is no need for a binding account to posit a perfect match in reference-fixing criteria. As a consequence, the binding model can accommodate a much wider range of ignorance, error, and direct logical disagreement than a matching model.

This flexibility about the substantive understanding required for direct logical relations, we submit, fits with a plausible epistemology of reference. From the point of view of a rational epistemic agent, one's higher-order opinions about the best epistemic methods for correcting one's first-order assumptions about the reference are no more infallible than the first-order assumptions themselves. The importance of stability of idiolect meanings and concepts has been emphasized by externalists like Putnam and Burge. Indeed, we take the real challenge to metasemantic internalism to derive from our epistemological commitment to *de jure* co-reference through open-ended inquiry and debate, rather than from simple conceivability arguments (Schroeter 2008).

The University of Melbourne

NOTES

1. To simplify the exposition, we'll assume that certain idiolect meanings map directly onto concepts: when one uses a context-neutral expression like "arthritis," one thereby expresses a context-neutral concept [arthritis].

2. This formulation of internalism is based on Chalmers and Jackson (2001). To avoid trivialization of the internalist position, the descriptions of the possible worlds must not (i) use the target word or concept, and (ii) include semantic or metasemantic information about the target word or concept. This internalist metasemantic framework presupposes that there is some base-level vocabulary for describing possible worlds that affords unproblematic epistemic access to objective properties—a claim that externalists would challenge (Stalnaker 2008). For a systematic elaboration and defense of this internalist commitment, see Chalmers (2012).

3. Proponents of a "deferential" model of concepts and idiolect meanings include Chomsky (1986); Peacocke (1992); Fodor (1994); Chalmers and Jackson (2001); Goldberg (2009).

4. For this line of argument, see Chalmers and Jackson (2001).

5. Influential advocates of this use of the 2-D semantic framework include Lewis (1994); Chalmers (1996, 2006); Jackson (1998); Chalmers and Jackson (2001). For an overview of different approaches to 2-D semantics, see Schroeter (2010).

6. Like more standard descriptivist analyses, a 2-D characterization of the semantic value of an expression must posit a base-level vocabulary that affords epistemically direct access to the fundamental properties that define a possible scenario considered as actual. Chalmers and Jackson suggest that epistemically possible scenarios can be specified in terms of a base-level vocabulary, PQTI, which includes descriptive terms for basic microphysical and phenomenal properties (P and Q), a "that's all"

clause (T), and an "I am here" self-locating clause (I). See Chalmers and Jackson (2001); Chalmers (2006, 2012), for details about the proposed epistemic 2-D framework.

7.   For a structurally equivalent division of the central tasks of a theory of concepts, see Peacocke's distinction between a theory of possession conditions and a determination theory (1992). As indicated above, we will take the questions of concepts and idiolect meanings to be equivalent in the cases we focus on.

8.   Burge (1979, 1982, 2010). See also Putnam (1970, 1973, 1975); Kripke (1980); Devitt (1981).

9.   Lewis (1970, 1979); Jackson (1998); Segal (2000); Chalmers (2002a); Jackson (2004).

10. Schroeter (2008, 2012); Schroeter and Schroeter (2014b). On pain of vicious regress, the mechanisms underlying this organizational structure cannot be mediated by explicit identity judgments of the form ⌜$a = b$⌝ or ⌜the referent of "a" = the referent of "b"⌝. The moral here is much the same as the one Lewis Carroll drew in the case of *modus ponens*: we must have some basic way of taking two thoughts to be co-referential that does not require an explicit identity judgment (Carroll 1895).

11. At the meta-level, in contrast, it's possible to cite undermining or reinforcing evidence for the claim that certain attitudes within a conscious episode do in fact co-refer. Our point is that, in object-level thought, the immediate appearance of sameness does not seem to depend for its justification on such meta-level evidence. It might seem that thoughts like those expressed by "war is war" are counterex-amples to this claim, since they can be justified by empirical evidence. However, it's important to keep in mind that our claim is targeted at the level of thought rather than language, and the token elements of thought expressed by the two tokens of "war" in this sort of non-literal usage will not strike the subject as pertaining *de jure* to the same topic. (Thanks to an anonymous referee for pressing us to clarify this point.) For further details about the appearance of *de jure* sameness and its role in a theory of concepts, see Schroeter (2012).

12. For a proposal in this spirit, see Chalmers (2011). Alternatively, a proponent of the matching model of concept identity can maintain that concepts are individuated in terms of reference-fixing criteria alone. In that case, a subject and her twin would share the same concept, but this fact wouldn't guarantee same-ness of reference. Concepts expressed by apparently context-neutral words like "water" or "arthritis" would thus be assimilated to the type of content expressed by indexicals like "here" and "now."

13. For a discussion of this approach, see Schroeter (2007, 2008); Schroeter and Schroeter (2014b). For alternative ways of accommodating the mismatch between apparent and real *de jure* sameness, see, for example, Burge (1998); Fine (2007).

14. We address some of these questions elsewhere (Schroeter and Schroeter 2014b; 2015).

15. Obviously, if one defers to a group that in turn takes one's own reference-fixer into account, this introduces circularity and no reference will be determined.

16. Of course, this is not to say that one can reject *all* substantive assumptions associated with a word at once while preserving the appearance of *de jure* sameness.

17. Even among co-conscious thoughts, direct binding relations seem explanatorily prior to a match in associated patterns of understanding. Any mechanism for keeping track of sameness must ultimately bottom out in some direct binding relations, on pain of a vicious explanatory regress. In particular, there must be some subpersonal mechanisms that directly bind together cognitive dispositions *at a time* as pertaining to a single topic, without relying on a match in the independent properties of those disposi-tions. But if a cognitive mechanism needs to invoke binding at some point, it's hard to see why this mechanism wouldn't be recruited in directly binding together two token "Hesperus" thoughts: Why posit a subpersonal mechanism that binds each token separately to a set of reference-fixing dispositions and then looks for a precise match in those dispositions before treating them as obviously and incontrovert-

ibly pertaining to the same topic? It's more economical to suppose that two token "Hesperus" thoughts are directly bound together within the very same bundle of attitudes and dispositions (Schroeter 2012).

REFERENCES

Burge, Tyler. 1979. "Individualism and the Mental," *Midwest Studies in Philosophy*, vol. 4, no. 1, pp. 73–121.

———. 1982. "Other Bodies," in *Thought and Object*, ed. Andrew Woodfield (Oxford, UK: Oxford University Press), pp. 97–120.

———. 1986. "Intellectual Norms and Foundations of Mind," *Journal of Philosophy*, vol. 83, no. 12, pp. 697–720.

———. 1988. "Individualism and Self-Knowledge," *Journal of Philosophy*, vol. 85, no. 11, pp. 649–663.

———. 1989. "Wherein Is Language Social?," in *Reflections on Chomsky*, ed. Alexander George (Oxford, UK: Blackwell), pp. 175–191.

———. 1998. "Memory and Self-Knowledge," in *Externalism and Self-Knowledge*, ed. Peter Ludlow and Norah Martin (Palo Alto, CA: CSLI Publications), pp. 351–371.

———. 2010. *Origins of Objectivity* (Oxford, UK: Oxford University Press).

Carroll, Lewis. 1895. "What the Tortoise Said to Achilles," *Mind*, vol. 4, no. 14, pp. 278–280.

Chalmers, David. 1996. *The Conscious Mind* (New York: Oxford University Press).

———. 2002a. "The Components of Content," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David Chalmers (Oxford, UK: Oxford University Press), pp. 608–633.

———. 2002b. "On Sense and Intension," *Philosophical Perspectives*, vol. 16, pp. 135–182.

———. 2006. "The Foundations of Two-Dimensional Semantics," in *Two-Dimensional Semantics: Foundations and Applications*, ed. Manuel Garcia-Carpintero and Josep Macià (Oxford, UK: Oxford University Press), pp. 55–140.

———. 2011. "Propositions and Attitude Reports: A Fregean Account," *Noûs*, vol. 45, no. 4, pp. 595–639.

———. 2012. *Constructing the World* (Oxford, UK: Oxford University Press).

Chalmers, David, and Frank Jackson. 2001. "Conceptual Analysis and Reductive Explanation," *Philosophical Review*, vol. 110, no. 3, pp. 315–361.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use* (New York: Praeger).

Cumming, Samuel. 2013. "From Coordination to Content," *Philosophers' Imprint*, vol. 13, no. 4, pp. 1–17.

Devitt, Michael. 1981. *Designation* (New York: Columbia University Press).

Dretske, Fred. 1988. *Explaining Behavior* (Cambridge, MA: Bradford/MIT Press).

Evans, Gareth. 1973. "The Causal Theory of Names," *Proceedings of the Aristotelian Society*, supp. vol. 47, pp. 187–208.

Fiengo, Robert, and Robert May. 2006. *De Lingua Belief* (Cambridge, MA: MIT Press).

Fine, Kit. 2007. *Semantic Relationalism* (Oxford, UK: Blackwell).

Fodor, Jerry. 1994. *The Elm and the Expert* (Cambridge, MA: MIT Press).

Goldberg, Sandord. 2009. "Experts, Semantic and Epistemic," *Noûs*, vol. 43, no. 4, pp. 581–598.

Heim, Irene. 1982. "The Semantics of Definite and Indefinite Noun Phrases," PhD thesis, University of Massachusetts.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford, UK: Oxford University Press).

———. 2004. "Why We Need A-Intensions," *Philosophical Studies*, vol. 118, nos. 1–2, pp. 257–277.

Kaplan, David. 1990. "Words," *Proceedings of the Aristotelian Society*, supp. vol. 64, pp. 93–119.

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

Lewis, David. 1970. "How to Define Theoretical Terms," *Journal of Philosophy*, vol. 67, no. 13, pp. 427–446.

———. 1979. "Attitudes *De Dicto* and *De Se*," *Philosophical Review*, vol. 88, no. 4, pp. 513–543.

———. 1994. "Reduction of Mind," in *A Companion to the Philosophy of Mind*, ed. Samuel Guttenplan (Oxford, UK: Blackwell), pp. 412–431.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories* (Cambridge, MA: MIT Press).

———. 1993. "White Queen Psychology: or, The Last Myth of the Given," in *White Queen Psychology and Other Essays*, ed. Ruth Garret Millikan (Cambridge, MA: MIT Press), pp. 279–363.

———. 2000. *On Clear and Confused Ideas* (Cambridge, UK: Cambridge University Press).

Peacocke, Christopher. 1992. *A Study of Concepts* (Cambridge, MA: MIT Press).

Perry, John. 1980. "A Problem about Continued Belief," *Pacific Philosophical Quarterly*, vol. 61, no. 4, pp. 317–332.

———. 2001. *Reference and Reflexivity* (Palo Alto, CA: CSLI Publications).

Pinillos, N. Ángel. 2011. "Coreference and Meaning," *Philosophical Studies*, vol. 154, no. 2, pp. 301–324.

Putnam, Hilary. 1970. "Is Semantics Possible?," in *Language, Belief and Metaphysics*, ed. Howard. E. Kiefer and Milton Karl Munitz (New York: State University of New York Press), pp. 50–63.

———. 1973. "Explanation and Reference," in *Conceptual Change*, ed. Glenn Pearce and Patrick Maynard (Boston, MA: D. Reidel), pp. 199–221.

———. 1975. "The Meaning of 'Meaning,'"in *Language, Mind, and Knowledge*, ed. Keith Gunderson, Minnesota Studies in the Philosophy of Science, Vol. 7 (Minneapolis, MN: University of Minnesota Press), pp. 131–193.

Schroeter, Laura. 2003. "Gruesome Diagonals," *Philosophers' Imprint*, vol. 3, no. 3, pp. 1–23.

———. 2004. "The Limits of Conceptual Analysis," *Pacific Philosophical Quarterly*, vol. 85, no. 4, pp. 425–453.

———. 2006. "Against *A Priori* Reductions," *Philosophical Quarterly*, vol. 56, no. 225, pp. 562–586.

———. 2007. "The Illusion of Transparency," *Australasian Journal of Philosophy*, vol. 85, no. 4, pp. 597–618.

———. 2008. "Why Be an Anti-Individualist?," *Philosophy and Phenomenological Research*, vol. 77, no. 1, pp. 105–141.

———. 2010. "Two-Dimensional Semantics," in *Stanford Encyclopedia of Philosophy* (Winter 2010 edition), ed. Edward M. Zalta. http://plato.stanford.edu/archives/win2010/entries/two-dimensional-semantics/.

———. 2012. "Bootstrapping Our Way to Samesaying," *Synthese*, vol. 189, no. 1, pp. 177–197.

———. 2014. "Scrutability and Epistemic Updating: Comments on Chalmers's *Constructing the World*," *Analysis Reviews*, vol. 74, no. 4, pp. 638–651.

Schroeter, Laura, and François Schroeter. 2009. "A Third Way in Metaethics," *Noûs*, vol. 43, no. 1, pp. 1–30.

———. 2014a. "Jazz Redux: A Reply to Möller," *Philosophical Studies*, vol. 170, no. 2, pp. 303–316.

———. 2014b. "Normative Concepts: A Connectedness Model," *Philosophers' Imprint*, vol. 14, no. 25, pp. 1–26.

———. 2015. "Rationalizing Self-Interpretation," in *Palgrave Handbook of Philosophical Methods*, ed. Christopher Daly (Basingstoke: Palgrave Macmillan), pp. 419–447.

Segal, Gabriel. 2000. *A Slim Book about Narrow Content* (Cambridge, MA: MIT Press).

Speaks, Jeff. 2010. "The Epistemic Two-Dimensionalism and the Epistemic Argument," *Australasian Journal of Philosophy*, vol. 88, no. 1, pp. 59–78.

Stalnaker, Robert. 1993. "Twin Earth Revisited," *Proceedings of the Aristotelian Society*, vol. 93, pp. 297–311.

———. 2004. "Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics," *Philosophical Studies*, vol. 118, no. 1–2, pp. 299–322.

———. 2008. *Our Knowledge of the Internal World* (Oxford, UK: Clarendon Press).

Yablo, Stephen. 2000. "Textbook Kripkeanism and the Open Texture of Concepts," *Pacific Philosophical Quarterly*, vol. 81, no. 1, pp. 98–122.