Legal Theory, **26** (2020), 62–99.
© The Author(s), 2020. Published by Cambridge University Press 0361-6843/20
doi:10.1017/S1352325220000063

A NEW INTERPRETIVIST METASEMANTICS FOR FUNDAMENTAL LEGAL DISAGREEMENTS

François Schroeter
Department of Philosophy, University of Melbourne

Laura Schroeter
Department of Philosophy, University of Melbourne

Kevin Toh*
Faculty of Laws, University College London

ABSTRACT
What does it take for lawyers and others to think or talk about the same legal topic—e.g., defamation, culpability? We argue that people are able to think or talk about the same topic not when they possess a matching substantive understanding of the topic, as traditional metasemantics says, but instead when their thoughts or utterances are related to each other in certain ways. And what determines the content of thoughts and utterances is what would best serve the core purposes of the representational practice within which the thought or utterance is located. In thus favoring a "relational model" in metasemantics, we share Ronald Dworkin's goal of explaining fundamental legal disagreements, and also his reliance on constructive interpretation. But what we delineate is a far more general and explanatorily resourceful metasemantics than what Dworkin articulated, which also bypasses some controversial implications for the nature of law that Dworkin alleged.

* Early drafts of this paper (with different titles) were presented at the Workshop on Law and Language at the University of Surrey School of Law in January 2017, the Workshop on Conceptual Analysis and the Future of Jurisprudence (as a part of the Edinburgh Legal Theory Festival) at the University of Edinburgh Law School in June 2018, and the UCL-Yale Symposium on the Future of Jurisprudence at University College London in June 2018. We thank those in the audience on these occasions for helpful questions and feedback. We also thank Eliot Michaelson for an instructive discussion and helpful advice.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

I. INTRODUCTION

A newspaper has reported that a certain local businessman has been providing the police and the district attorney's office with some helpful information for their investigation of a reputed mafia boss. The businessman's lawyer, Brooke, has promptly filed a defamation suit against the newspaper. Fred, the newspaper's lawyer, has responded that only a publication that tends to lower a person's reputation among "the right-thinking people" can amount to defamation. Since helping the police and prosecutors with their investigations and prosecutions is what right-thinking, law-abiding citizens expect of one another, Fred argued, the newspaper's reporting could not amount to any defamation of Brooke's client. Brooke concedes that in many defamation cases, perhaps even most, the opinions of the law-abiding citizens are decisive. But that is so, she argues, because the harms to the plaintiffs in those cases are best measured in terms of the changes in the opinions of "the right-thinking people". But when a plaintiff belongs to a particular kind of social milieu, as her client does, the reputation of being a "snitch" is the most harmful and dreaded one a person could sustain. It would lead to complete social ostracism, she elaborates, and may even threaten one's life. "Snitches end up in ditches", as the saying goes.

Brooke and Fred clearly differ in their "theories" about defamation. But also quite clearly they are both talking and thinking about the same topic—namely, defamation, or defamation in their jurisdiction. It may turn out that one is a better lawyer than the other, and will eventually prevail in court. But any difference in their *legal expertise* does not entail a difference in their ability to think and talk about defamation—i.e., in their *conceptual competence*.

Roughly and broadly put, one of our questions is: In virtue of what do lawyers think and talk about the same legal topic, such as defamation? The foregoing considerations suggest that lawyers do not manage to think and talk about some legal topic in virtue of their shared theory or conception of that topic. Brooke and Fred after all do not agree on what it takes for some publication to amount to defamation; and yet, they manage to think and talk about the same tort of defamation in their jurisdiction. So how do they manage to accomplish that?

A tempting move at this point is to trace the "competence conditions" not to the shared conception of defamation, but something more fundamental that Brooke and Fred share. Perhaps they share a conception of tort (in their jurisdiction), or even more fundamentally a conception of law. But one also suspects that analogues of the foregoing considerations would apply at these deeper levels as well. Brooke might be of the view that people have interests in avoiding certain kinds of harm, and that the purpose of the tort law in her legal system is to afford compensation when such interests are not outweighed by societal interests in allowing defendants to engage in certain kinds of activities, including certain kinds of speech; whereas

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

Fred might subscribe to a competing conception of tort law according to which individuals have certain rights to bodily and reputational integrity, and the purpose of tort law is to provide avenues of redress when those rights are infringed.[1] Even more fundamentally, Brooke may be of the view that legal systems as such have the proper function of preventing certain kinds of harms and affording certain kinds of benefits; whereas Fred may be of the view that legal systems are by nature deontic systems of artificial rights and obligations that need not be functions of considerations of human welfare. Even in the event of such very fundamental legal disagreements, it appears, Brooke and Fred could manage to think and talk about the same legal topic—whether defamation, tort, or even law itself. Our first question is: How do they manage to do so?

Our second question would be prompted by our answer to that first question. That second question is: What does our answer to the first question suggest about the nature of law?

II. SOME PRECEDENTS

Readers who are familiar with the contemporary legal philosophical literature would certainly see that our questions are not entirely new. The disagreements of the sort that we discussed in the preceding section, or at least the most fundamental of them, are those that Ronald Dworkin called "theoretical disagreements" in law in his seminal discussion in Chapter 1 of *Law's Empire* (1986). Dworkin argued that theoretical disagreements are actual and ubiquitous in real and flourishing legal systems, and that then-leading approaches to thinking about the nature of law lacked resources to explain them.

The "tempting move" that we discussed in the penultimate paragraph of the preceding section is one that Dworkin attributed to his predecessors, most notably the H.L.A. Hart of *The Concept of Law* (1961/1994). In a community with a genuine and working legal system, Hart opined, the bulk of the officials of the system accept in common what Hart called "the rule of recognition", which is the norm that enumerates the most fundamental criteria of legal validity in that system. It is then this shared conception of legality in that system that enables competent lawyers to think and discuss, and sometimes disagree about, legal matters in the relevant community. Of course, lawyers may disagree with each other about many legal matters. But the thought is that they must share a certain limited set of assumptions about legality in their legal system, and these assumptions serve as a kind of topic-specific template for recognizing what it takes to be a law in their legal system. This limited set of assumptions, the template, is a conception of the rule of recognition of the relevant legal system. Lawyers and others think

1. For the distinction between these two conceptions of tort law, and an elaboration and defense of the latter conception, *see* Stevens (2007).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

and talk about the law of the same legal system insofar as they share that template or conception. If two lawyers or discussants accepted different conceptions of the rule of recognition of their legal system, then they in effect would be talking about the laws of different jurisdictions, and would consequently talk at cross-purposes.[2]

Dworkin appealed to the phenomenon of theoretical disagreements to discredit this Hartian picture. Pointing to some real and imagined legal disputes, he argued that discussants, especially judges, need not agree on a conception of the rule of recognition of their legal system to have a genuine discussion and even disagreements about the law of their community. Hart in effect, according to Dworkin, required too much for people to be thinking and talking about the law of their community, and thereby limited too much the kinds of genuine legal disagreements—as opposed to disagreements about what the law should be on the one hand, and merely apparent disagreements on the other—that lawyers and others could have with each other. Dworkin went on to devise an alternative conception of the nature of law. And a central component of Dworkin's theory of the nature of law was his view of the considerations in virtue of which two lawyers, or any discussants, could be deemed to be thinking or talking about the same topic.

The precise shape of Dworkin's view on this last issue has never been entirely clear, and it has also shifted over the years. Originally, Dworkin cast the targets of his criticisms as offering semantic theories of law. In effect, he read the various criteria of legality that the different theories discussed, and the subordinate laws that are valid according to such criteria, as competing conceptions of the meaning of the word "law" and of its extension, respectively. Dworkin argued that semantic theories, so conceived, are incapable of explaining theoretical disagreements in law, whereas what he offered in their stead, the so-called interpretive theories of law, could and hence should be preferred. Initially, many thought this aspect of Dworkin's exposition puzzling and even perverse, something to be ignored out of charity, since Dworkin's predecessors had not offered semantic theories of the sort that he was picturing.

Subsequently, Nicos Stavropoulos (1996) offered a revisionist reading of Dworkin's argument that Dworkin himself eventually came to adopt. According to Stavropoulos, what Dworkin was getting at was not best conceived as a contest between semantic theories and interpretive theories of law, but as one between legal theories that employ criterial semantics and those that employ interpretive semantics. In other words, both kinds of theories of law have semantic commitments. The problem with older legal theories, according to Stavropoulos's version of Dworkin's argument, is not that they employ semantics, but that they rely on a kind of semantics that

2. For a conception of Hart's legal theory that differs from the influential conception summarized in the text, and that arguably is not susceptible to Dworkin's diagnosis and criticisms, *see* Toh (2005). *See also* Toh (2011 & 2019).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

lacks resources to account for theoretical disagreements in law. By the time of his swansong, *Justice for Hedgehogs* (2011), Dworkin clearly came to embrace this take on his 1986 argument. Pushing Stavropoulos's line of thinking, Dworkin there adopts the view that there are two kinds of concepts. He says near the very beginning:

> We must ask: When do people share a concept so that their agreements and disagreements are genuine? We share some concepts because we agree, except in cases we all regard as borderline, about what criteria to use in identifying examples. . . .
>
> We must . . . recognize that we share some of our concepts, including the political concepts, in a different way: they function for us as interpretive concepts. We share them because we share social practices and experiences in which these concepts figure. We take the concepts to describe values, but we disagree, sometimes to a marked degree, about what these values are and how they should be expressed. We disagree because we interpret the practices we share rather differently: we hold somewhat different theories about which values best justify what we accept as central or paradigm features of that practice. That structure makes our conceptual disagreements about liberty, equality, and the rest genuine. (2011, p. 6)

According to Dworkin, his predecessors made the mistake of thinking that LAW is a criterial concept, when in fact it is an interpretive concept.[3]

We find what Dworkin offers as the alternative in the second paragraph very suggestive and even congenial. Yet, there are a number of problems with Dworkin's proposal that we want to correct. The first and most obvious problem has to do with the programmatic nature, the sheer amount of hand-waving, that the proposal involves. Dworkin's stated view is that we are able to think and talk about a same topic "because we share social practices and experiences in which . . . [the relevant] concepts figure". We would like to be much clearer and more specific than Dworkin as to what if anything we share with each other when we manage to think and talk about same topics, and also about how exactly what we share determines or picks out the relevant topics or semantic values. Second, while Dworkin is not opining merely about specifically legal concepts but instead about normative and evaluative concepts in general, he does not explain how this proposal about such concepts fits into a general account of concepts. Dworkin conceives normative and evaluative concepts as interpretive concepts, and contrasts them with criterial concepts. Eventually, of course, a proposal like Dworkin's needs to work out how different kinds of concepts can be used or deployed together in our thinking and talking. We propose below a general theory of concepts that bypasses the kind of distinction that Dworkin is committed to, and the difficulties it incurs. Third, although Dworkin initially portrayed his arguments as discrediting both legal positivist

3. In this paper, we will use SMALL CAPS to denote concepts.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

and natural law theories, he eventually took on the view that there is a short distance between his interpretivist conception of normative or evaluative concepts on the one hand and a natural law conception of the nature of law on the other. We are skeptical, and we argue below that a plausible theory of concepts along the lines that Dworkin suggests does not favor one or the other traditional conception of the nature of law. Fourth and finally, Dworkin's interpretivist conception of normative and evaluative concepts is meant to enable him to steer clear of metaethical or meta-normative commitments. Again, we very much doubt Dworkin's thinking on this issue, and we argue below that a plausible theory of concepts does not quite have the "quietist" upshots that Dworkin favored and sought to vindicate.

What we propose and motivate in subsequent pages is a conception of what it takes for lawyers and others to think and talk about the same legal topic—say, defamation in a jurisdiction—that does not require them to share a conception of that topic, or some more fundamental conception of legality in that jurisdiction. Our motivation, as was Dworkin's, is to allow for radical legal disagreements of the sort that Dworkin called "theoretical disagreements". And like Dworkin, we will be carving out a central role for interpretation, and more specifically for the kind of interpretation that Dworkin calls "constructive interpretation". At the same time, what we propose, in our view, does not suffer from the kinds of problems that we have just enumerated that Dworkin's proposal suffers from.[4] What we end up with, we believe, is a new interpretivist conception of what makes fundamental legal disagreements possible.

III. TWO MODELS OF CONCEPTUAL COMPETENCE

In the course of this and the next two sections, we will set out two distinct models of what it takes for lawyers and judges, or people more generally, to think and talk about the same topic, legal or otherwise. One of the models is the traditional view, along the lines of what Dworkin has attributed to Hart and criticized. The other is what Dworkin's own view *approximated*, what we think Dworkin *should have* devised in place of the traditional model. Let us call these two models "the matching model" and "the relational model", respectively, for reasons that should become obvious presently. In this section, we appeal to an analogy and an example to give the reader some intuitive feel for the difference between the two models. We will be able to flesh out the models more fully two sections hence, after doing some needed conceptual housekeeping in the next section.

4. Although we consider what we delineate below much clearer and more detailed in many respects than Dworkin's account, in a short paper like this, we ourselves will not be able to acquit ourselves entirely of the possible charge of being programmatic. Still, we trust, the reader will get a much better idea of what our subsequent steps would look like than he would with Dworkin's account. We also hope to follow up this paper with further work that will trace out some of those steps.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

First, the analogy. Consider for a moment the issue of personal identity. In the pertinent philosophical literature, there have been, broadly speaking, two approaches to explaining personal identity—i.e., what it takes to be the same person over time. The traditional approach explains personal identity in terms of some common properties that different person-stages share. Two person-stages are stages of the same person, according to this approach, just in case they are bound by the same immaterial soul, the same core set of memories, or some other common properties. An alternative approach explains personal identity in irreducibly relational terms—that is, in terms of relations among person-stages that cannot be reduced to some properties that the person-stages have in common. According to this latter approach, two person-stages are stages of the same person not because they share some common properties, but because their psychological states are causally related to or connected with each other in certain ways. These two approaches to personal identity have analogues, respectively, in two approaches to thinking about people's ability to think and talk about the same topic, or the two models that we are calling "the matching model" and "the relational model".

The matching model explains people's ability to think and talk about the same topic by way of a common or matching substantive understanding of the topic that each possesses. Alternatively, the relational model explains the same ability by appealing to the causal relations or connections among the people's thoughts or utterances. Now, the following example of a conversational exchange is meant to illustrate the difference between the two models, and at the same time provide some motivation for the relational model:

(1) A: "Ruth Crawford Seeger was the most important American composer of the twentieth century."
B: "No, she wasn't!"
C: "Sorry, but who is she?"

In this exchange, it appears quite obvious to both the participants and us the observers that the three people are discussing a single common topic—namely, Ruth Crawford Seeger. But clearly not all parties share the same substantive understanding of the topic in discussion. Presumably, A and B have rich (albeit probably divergent) understandings of the characteristics of the American composer picked out by the name "Ruth Crawford Seeger"; whereas C may know nothing more than that Crawford Seeger is the person (if any) to whom A and B have just referred. The apparent sameness of topic here therefore could not be grounded in any matching (or even similar) substantive understanding of the topic in discussion. Instead, it seems plausible to think, the apparent sameness of topic is secured by the fact that the three discussants' utterances, as well as their associated thoughts,

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

are linked together in a certain way. If B's and C's respective sentences had occurred within quite different chains of conversation, there would have been no obvious sameness of topic, it seems. These considerations favor the relational model over the matching model.

This last example is one involving discourse anaphora, in which different utterances within the conversation are linked together by people's pre-reflective parsing mechanisms in such a way that a proper name and subsequently used pronouns appear obviously to pertain to the same topic.[5] We want to generalize and embellish this relational account. For one thing, we want to account for talk about the same topic even when the expressions used are not obvious anaphors. Second, we want to explain how people think as well as talk about the same topic. To make headway in this project, we need to do some conceptual housekeeping first. To that immediate task, we now turn.

IV. SOME HOUSEKEEPING

Up until now, we have been speaking rather casually and loosely about topics and concepts, and have been going back and forth between the two notions as we broached our discussion. In our subsequent discussion, we would like to be more explicit and scrupulous about the relation between these two notions, as well as with the related notion of linguistic meaning, or at least among these notions as we will be regimenting them shortly.

Notice to begin that two people may think or talk about the same topic without realizing so, as the following example shows:

(2) A: "Hesperus is in the Western sky."
B: "Phosphorus is in the Eastern sky."

Both "Hesperus" and "Phosphorus" refer to the planet Venus, and it follows that the two people are thinking or talking about the same topic in this example. But they may fail to realize it. In fact, a single individual may hold these two thoughts, or make these two utterances, without realizing that he is thinking or talking about the same topic. To borrow a way of regimenting the relation between topics and concepts that comes down to us ultimately from Gottlob Frege (1892),

(3) If two token elements of thought do not seem, from the perspective of the single individual who entertains them, to pertain obviously and unquestionably to the same *topic*, then they express different *concepts*.

5. Discourse anaphora raises a number of difficult semantic and philosophical puzzles. For an overview of these issues and alternative semantic theories, *see* King & Lewis (2018). The approach to concepts we suggest below is related to discourse representation theory, pioneered by Hans Kamp (1981) and Irene Heim (1982).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

Here, the notion of a concept is tied to individual epistemic perspectives on sameness of topic. Although an individual's "Hesperus" and "Phosphorus" thoughts concern the very same topic (namely, Venus), this sameness may not be obvious and unquestionable from his subjective perspective. It would then be compatible with minimal standards of rationality for him to doubt that Hesperus is Phosphorus, or to believe that Hesperus has some properties that Phosphorus lacks. According to Frege, such divergent thoughts toward the same topic indicate that these thoughts involve different *ways of thinking* about the same topic, which is what we are calling "concepts".[6]

Similarly, when a person deploys a single concept on multiple occasions, it seems to him as if he is *guaranteed* to be thinking about the same topic on those occasions. Consider a bit of conscious reasoning:

(4) C: "Hesperus is in the Western sky."
C: "If Hesperus is in the Western sky, it won't be obscured by cloud tonight."
C: "So Hesperus won't be obscured by cloud tonight."

From the thinker's perspective, it seems obvious and unquestionable that there is just one topic in question throughout. Moreover, semantic theorists agree that this sort of reasoning does not just *appear* logically valid from the thinker's limited perspective, but also that it really *is* valid, at least in normal circumstances. A person's deploying the same concept ensures both the subjective *appearance* of guaranteed sameness of topic and the *reality* of sameness of topic. The same can be said when different people share a concept as they think together about a topic. For instance, the reasoning instantiated in (4) could occur in a conversation among three different individuals. Here too, the subjective appearance of guaranteed sameness of topic would be a reliable guide to the reality.[7]

Such an appearance of *guaranteed* sameness of topic can be contrasted with an appearance of merely *de facto* sameness of topic that is characteristic of Hesperus/Phosphorus cases of the sort described above. To put it

6. According to Gibbard (2003, ch. 2), G.E. Moore's "open question argument" invokes a similar constraint on concept identity. The fact that the question "It's pleasant, but is it good?" strikes us as open suffices to establish that two different concepts figure in that question.

7. It is worth emphasizing that the appearance of guaranteed sameness can in principle be mistaken—even in the case of an individual thinker. For instance, a person might mistakenly think of two identical twins as a single individual. In that case, the thinker accumulates a body of information over time about what *seems* to him obviously and unquestionably like a single individual. However, once he is apprised of the fact that he's been interacting with two people rather than one, it's natural for him to reject the earlier appearance of guaranteed sameness of topic as mistaken. From this more informed perspective, it seems that *those* beliefs were about one twin, while *these* were about the other. For a defense of the corrigibility of appearances of guaranteed sameness and how this impacts concept identity, *see* Burge (1998) and Schroeter (2007 & 2008). Fortunately, such cases of unwitting conflation are rare, and the appearance of guaranteed sameness is a generally reliable guide to the facts about sameness of reference.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

another way, two people's sharing a concept, or one person's deployment of one concept on multiple occasions, is more demanding than sameness of topic alone. Sharing or using the same concept also ensures that people or a person over time thinks about that topic in the same *way*. The intuitive idea here is that conceptual competence makes sameness of topic subjectively accessible to persons themselves, independently of any empirical inquiry. And as *concepts* furnish apparent guarantees of sameness of topic *in thought*, linguistic *meanings* furnish apparent guarantees of sameness of topic *in talk*.[8]

V. TWO MODELS REDUX

With the drudgery of such housekeeping out of the way, we are in a position to convey fuller conceptions than in Section III of the two models of our ability to think and talk about the same topic. What we are after is not an explanation of how one or more people merely *happen to* or *coincidentally* think or talk about a single topic over time or in the course of a conversation, as they do in Hesperus/Phosphorus cases. Thinking or talking about a single topic over time, or among two or more people, is something that we implicitly intend and (oftentimes) accomplish. What we are after is an explanation of what devices or infrastructure we rely on for such an accomplishment. In other words, what we are interested in is not so much mere *de facto* sameness of topic, but instead the subjectively obvious *guarantee* of sameness of topic that concepts and meanings are meant to furnish us.

Although thinking or talking about a single topic over time or between two or more people is an accomplishment, it amounts merely to getting an entry ticket into the arena of investigating or deliberating about the topic in question. To inquire and learn about a topic, it is essential to keep track of different thoughts and utterances about it; but doing so is a mere preliminary to a genuine cognitive achievement of learning about the topic. It makes sense then to conceive the ability to keep track of various thoughts and utterances by deploying a concept and an associated meaning as a *competence*. Many people are competent in this sense with respect to a topic, and such competence may necessarily involve an ability to form some new true beliefs or justified commitments about the topic. Still, such competence falls short of *expertise* or *virtuosity* that would be necessary for people to reliably refine their current rough understanding of the topic in the direction of a better, more accurate understanding of that topic.

In the preceding section, we explicitly borrowed from Frege, or more accurately the tradition that goes back to him, to connect concepts to

8. Concepts and context-neutral linguistic meanings are systematically aligned, though correspondence is imperfect because not all words express concepts, and not all of a person's concepts are associated with words. For more on the appearance of guaranteed sameness, *see* Schroeter (2012). For further details on this approach to concept individuation applied to normative concepts, *see* Schroeter & Schroeter (2014).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

individual epistemic perspectives on sameness of topic. Frege not only gave us this prompt for thinking about concepts, but also proposed a particular model of concepts—that is, a particular account of when two or more token elements of thought appear obviously and unquestionably to pertain to the same topic, as judged from a single subjective perspective. What is proposed is in effect the matching model of conceptual competence, according to which two thoughts deploy the same concept, or two utterances express the same meaning, and thereby generate an apparent guarantee of sameness of topic, if the two thoughts or utterances are associated with the very same conception of the topic in question. For example, Aristotle and Lavoisier shared the same concept WATER if their thoughts were associated with the very same (implicit) conception of water—say, as the clear potable liquid that actually fills rivers, lakes, and seas. And anyone would be deemed competent to think and talk about water, according to this model, if he subscribes to the same substantive conception of water.[9]

Our example in Section III of discourse anaphora involving Ruth Crawford Seeger, as well as the example involving defamation that we began this paper with, cast doubt on this Fregean matching model of concept possession. In effect, the matching model makes the entry ticket too pricey. The aforementioned examples suggest that people should be able to secure a common topic in their thoughts and utterances, and be able to keep track of a single topic in their varied and divergent thinking and talking, even when they do not share some substantive conception of the topic in question. As a matter of fact, the possibility of open-ended inquiry and debate about a topic seems to depend on people's ability to keep track of it without having to possess or share any particular conception of the topic. Perhaps Aristotle and Lavoisier did share some common substantive understanding or conception of water; and perhaps anyone who thinks or talks about water is likely to, though not necessarily, share the same. But people thinking or talking together are unlikely to have quite matching conceptions or understandings of matters like wombats, corrosion, roof shingles, attorney-client privilege, callowness, shoegaze chord patterns, etc.; and such lack is unlikely to hamper their thinking and talking together about any such topic and possibly learning from each other. And the same could be said about an individual person who learns more or forgets much about any such topic over time. The relational model of conceptual competence, which we favor over the matching model, opts for the position that the entry ticket to the arena of investigating or deliberating about any topic is afforded by having the right connections. The cognitive devices of concept and meaning that we use to keep track of thoughts or utterances

9. We include in the matching model both traditional descriptivist theories like those advocated by Frege (1892) and Russell (1911), and sophisticated neo-Fregean theories of concepts and meanings that can be found in, *e.g.*, Lewis (1970), Peacocke (1992), Jackson (1998), and Chalmers (2006). In the case of the latter group of theories, a matching conception is a necessary condition for sameness of concept, but not always a sufficient one.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

about a topic are not marked by any common conception or understanding of the relevant topic, but by being causally related to each other in certain ways.

It is not that we lose nothing by opting for the relational model over the matching model. Given that the matching model does not require any causal connections for thoughts and utterances to deploy the same concepts, it allows for concepts to be shared across causally isolated communities. All that is required, according to the model, is that people's thoughts and utterances be associated with the very same conception of the topic in question. On any relational model, in contrast, the requirement of causal connectedness rules out the possibility of causally isolated persons sharing the same concept.[10] We believe, however, that this cost is more than sufficiently offset by the gain in the relational model's portrayal of concepts as shareable by people without their having precise matches in associated conceptions or understandings. Like a relational account of personal identity, a relational account of concept identity allows for much more variation in associated understandings than any traditional matching account. We take this to be a decisive advantage of the relational approach. The matching model's inflexibility about what is required for apparent guarantee of sameness of topic clashes with our experience as thinkers. If concepts are to remain stable through open-ended inquiry and debate, their identity cannot depend on a precise match in associated conceptions. For example, someone could begin thinking or talking about wombats, while assuming that wombats are placental mammals, and only later come to realize that they are marsupials. One may even begin thinking or talking about water, while assuming that water is one of the most basic elements of nature, and only later come to realize that water consists of even more basic elements hydrogen and oxygen. Indeed, for any topic, it seems, any particular assumption, even one consisting of deep-seated implicit cognitive or motivational dispositions, could be dropped without threatening the appearance of guaranteed sameness of topic.[11]

In our view, however, the fact that no particular assumption or belief is a nonnegotiable part of the entry price does not mean that causal connections among token thoughts or utterances can be the sole determinants of concept identity that in turn provides apparent guarantee of sameness of topic. Sometimes, the requisite causal connections are there, but the substantive understandings or conceptions held by different people, or by the same person at different times, diverge so radically as to preclude any interpretation that vindicates the apparent guarantee of sameness of topic. That

10. Some have argued that this is a decisive point against the idea that ethical or normative concepts can represent determinate properties. *See, e.g.*, Hare (1952, pp. 148–150; 1986); Horgan & Timmons (1992). For some responses, *see* Dowell (2015); Schroeter & Schroeter (2016); Millikan (2017, pp. 95–96).

11. For recent defenses of this view, *see, e.g.*, Williamson (2006); Schroeter & Schroeter (2009).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

would be the case, for example, if the conversational exchange of (1) were to continue as follows:

(5) A: "Ruth Crawford Seeger was the most important American composer of the twentieth century."

B: "No, she wasn't!"

C: "Sorry, but who is she?"

D: "My grandmother took piano lessons from her."

E: "My uncle saw her at Café Bizarre when she was playing the drums for the Velvet Underground."

The fifth speaker is clearly an outlier in this exchange. Although his utterance, and the thought it expresses, stand in the same anaphoric relations with the rest of the utterances and thoughts in the exchange as the others, the radically divergent substantive understanding associated with his utterances precludes co-reference with the other four. Unlike the other participants in the exchange, he is not thinking about Crawford Seeger the composer, but about a different woman altogether—namely, Moe Tucker.

To accommodate the complication thus exemplified, our approach proposes two core conditions for competence with the same concept.

(6) In order to deploy the same concept,

(a) two utterances (and the thoughts they express) must be causally related in an appropriate way (*connectedness condition*); and

(b) the substantive understanding directly associated with those utterances (and the thoughts they express) must not diverge so radically as to preclude a univocal interpretation (*congruence condition*).

Consider a case of disagreement over which actions are morally right. As many theorists have noted, it seems possible to genuinely disagree about the moral status of actions that seem to most of us to be obvious and central cases of moral rightness, like helping Singer's (1972) drowning toddler. To argue that helping the toddler is not morally right does not undermine a person's ability to have a genuine disagreement with the majority, deploying the very same concept MORALLY RIGHT. Similarly, we can genuinely disagree about central moral principles, such as the principle that cruelty is morally wrong, without failing to share the very concept of MORALLY RIGHT. As long as there is sufficient overlap in *other* aspects of the contrarian's understanding of moral properties, his opinions will intuitively strike us as obviously pertaining to the same moral topic as our own. As the example exchange of (4) was meant to illustrate, this impression is defeasible and would be defeated if the contrarian's commitments were sufficiently aberrant across the board. If he were to hold, for example, that counting blades

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

of grass is the only morally right action, and remained impervious to all arguments about human welfare and interests, then we would likely conclude that he is not competent with our concept MORALLY RIGHT.[12] But any particular attitude or commitment seems optional, as long as the contrarian's understanding is sufficiently congruent overall with that of others. Thus adjusted, relationally specified competence conditions, we believe, afford a flexibility that would allow any sorts of open-ended debates that we would be interested in accommodating.[13]

VI. CONCEPTS AS MENTAL FILES

The two conditions for conceptual competence we just introduced for the relational model are highly schematic, and can be fleshed out in a myriad of different ways. In this section, we will focus on the connectedness condition, which requires that utterances be "causally related in the appropriate way". Just what are the causal relations that could ground a plausible relational theory of conceptual competence?

Let us start with a simple case of conceptual competence without language. Imagine that Fred, our newspaper lawyer, takes the tram to work every day and he notices a flamboyant woman who regularly gets on the tram a few stops down from him. Each day, she sports some fascinating and outrageous outfit, and Fred starts to wonder about her background and speculates about her line of work. Where could she possibly be going in that getup? She must be some sort of artist . . . or maybe a publicist? Or maybe she is just a highly eccentric trust-funder? As Fred continues to keep track of the woman, amassing new information, speculations, and attitudes about her, Fred is relying on a stable concept of her—a single *way* of thinking about her that presents all these thoughts as obviously pertaining to the same person. Notice that this ability to keep track of a person and amass information and attitudes about her does not depend on having any *name* for the person in question. The ability rests on basic cognitive capacities for binding information and attitudes together at a time as pertaining to the same topic, and for keeping track of that bundle of information and attitudes over time as one adds to or subtracts from the bundle.

12. For an influential defense of this position, *see* Foot (1959). For an account of competence with normative concepts that is at the more formal end of the spectrum, *see* Gibbard (1992 & 2003).

13. Here we are following the philosophical tradition in taking "genuine disagreement" to involve sharing the same concept. The rationale for this move is that sameness of concept grounds direct logical consistency and inconsistency. Sharing the same concept not only guarantees co-reference but explains why co-reference is subjectively obvious. These are the hallmarks of moral disagreement that theorists have sought to explain. However, there are alternative conceptions of "genuine disagreement" that do not posit a shared concept. Plunkett & Sundell (2013a & 2013b), for instance, introduce a broader conception of "genuine disagreement" that does not involve relations of logical consistency and inconsistency. We are setting aside this nonstandard conception of genuine disagreement in this paper.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

Many theorists have appealed to the metaphor of a *mental filing system* to characterize the subpersonal cognitive mechanisms that ground these abilities to bind and segregate informational states as obviously pertaining to the same topic.[14]

The point we want to emphasize here is that the ability to keep track of a topic at a time and over time relies on quasi-anaphoric causal relations among our own mental states. As Fred sits on the tram idly speculating about the woman's motivations and professional life, each thought in the sequence seems obviously to pertain to the same topic as the immediately preceding ones. Moreover, it does not seem like the appearance of obvious sameness of topic is due to some sort of resemblance in the intrinsic characters of these thoughts considered independently. After all, Fred might idly imagine that the woman has an identical twin somewhere in the city, dressing in identical ways.[15] What makes some of Fred's imaginative states about *this woman* in front of him, and others about her hypothetical twin sister, seems to hinge entirely on how they are causally related to Fred's perceptions of the woman sitting two rows away from him right now. In effect, Fred's subpersonal cognitive mechanisms quasi-anaphorically bind one pretense state to an existing bundle of perceptual states, while the other pretense state might launch a new and distinct bundle of attitudes.

Similar observations can be made about obvious sameness of topic over time. Many years later, Fred might remember the woman on the tram and wonder what has become of her. What make Fred's current curiosity one about the very same woman as his attitudes from his past encounters in trams are quasi-anaphoric causal connections linking his current state to his persisting memories, which are in turn causally derived from his original perceptual states. Without these causal links, Fred's question about a woman with certain past characteristics could very well be about a hypothetical twin rather than about the woman he encountered many years ago. The causal links themselves seem to be crucial to the obvious sameness of topic characteristic of a single concept—i.e., a single way of thinking about a topic.

Now, let us consider how language allows a person to coordinate with others. Fred might start telling his friends about the mystery woman on the tram, bringing them into his game of speculation and storytelling about her. To make things easier, Fred gives her a name, "Fabulous Fiona". The initial process of acquiring competence with the term will look a lot like

14. The mental file metaphor is a popular one, and it has been developed in different ways. *See, e.g.*, Evans (1973 & 1982); Strawson (1974); Perry (1980 & 2001); Forbes (1989); Recanati (1993, 2012 & 2016); Millikan (2000); Lawlor (2001); Schroeter (2008 & 2013); Jeshion (2010); Dickie (2010). The crucial points for our purposes are: (i) that mental files have the function of binding bundles of information *as* pertaining to a single topic; and (ii) that mental files are individuated by the binding mechanism itself rather than by the bound information.

15. As it happens, not a completely frivolous thought given the precedent of Marian and Vivian Brown in San Francisco.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

the discourse anaphora case in Section III. Fred's friends will acquire new attitudes that they take to pertain obviously to the same person (real or fictional) that Fred has in mind. This appearance of sameness would not be grounded in any resemblance between their respective conceptions of Fiona and Fred's. For their imaginary reconstructions of Fiona may actually be very different from Fred's perceptual memories of the woman in question. What would ensure sameness of topic seems to be a quasi-anaphoric binding mechanism. Fred's friends will proceed to form their own bundles of attitudes, including beliefs, speculations, and even fantasies, which they take to pertain obviously to the same topic as those of everyone else who has been initiated into Fred's game. Eventually, the legend of Fab Fi might take on a life of its own, and the links to the original sightings of the woman on the tram may play very little or no role in regulating the ways that the fantasies play out. Nonetheless, each person who has acquired the name will take others' use of the name to obviously pertain to the same topic as his own. In effect, different people's mental files would become linked by chains of quasi-anaphoric causal relations, and thereby amount to what could be deemed a *shared representational tradition.*[16]

Many theorists have argued for something like this relational model of intrapersonal and interpersonal binding relations in explaining our ways of thinking and talking about individuals, and perhaps about natural kinds like gold and water. Our proposal is to apply the relational approach across the board to all stable concepts, regardless of the ontological category of the topics or subject matters picked out by concepts. In particular, such a relational approach would help us to think about our ability to keep track of normative and legal topics, despite variation in substantive understandings over time and between persons, in ways that are more satisfying than what has been available to us in the legal philosophical literature that has been prompted largely by Dworkin's discussions of the issue.

Consider the term "legal", for example. At the level of thought, both laymen and experts are likely to have a wide variety of different attitudes about the topic in question—e.g., about which actions are illegal in one's community, how the legal status of certain action-types varies from jurisdiction to jurisdiction, how some law can be changed through legislation and enforced through various institutional means, what kinds of dispositions to obey the law are called for, what kinds of legal restrictions warrant resentment and reform, etc. When attitudes like these occur together within the same stream of thought, a person treats them as obviously pertaining to the same topic. Roughly, legality is a normative property or status that certain action-types have relative to a regulatory jurisdiction, which property or status can be modified, and the disputes about which can be adjudicated, through certain kinds of institutional activities. But each person's

16. *See* Schroeter & Schroeter (2014) for a more precise characterization of representational traditions.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

understanding of legality changes or evolves over time, and virtually any particular attitude could be replaced, revised, or rejected without jeopardizing this obvious sameness of topic over time. And there is likely even more variation in the substantive conceptions held by different people. There just does not appear to be any pattern of substantive understanding that is mandatory or nonnegotiable for obvious sameness of topic. Just as in thinking about individuals like Crawford Seeger, therefore, it is plausible to posit quasi-anaphoric causal connections that underwrite this apparent guaranteed sameness of topic in the case of properties like legality. The reason it seems obvious that we are all talking about the same topic, despite our disagreements over the nature of legality, is that we all belong to the same representational tradition.

Demarcation of representational traditions is a complicated matter. At the intrapersonal level, individual thoughts are bound together at a time and over time so as to generate the appearance of guaranteed sameness of topic by the workings of automatic cognitive binding mechanisms. What results is something like an internal mental filing system, in which each bound bundle of thoughts amounts to a mental file. Some of the mental files held by different individuals are then systematically linked together via our linguistic parsing mechanisms to be deemed as obviously about some single topic. A person automatically hears others' use of words as expressing a thought pertaining to the same topic as one of one's own mental files.[17] Shared representational traditions are then demarcated by chains of these sorts of causal relations linking token thoughts.

The resulting representational traditions will often extend to virtually the whole linguistic community—as is probably the case with "legal". But it may comprise only a discrete subcommunity—as with our imagined case of "Fabulous Fiona", or with a term of art like "the rule against perpetuities". And sometimes there can be two distinct traditions associated with the same word, as has been the case with "corn". British speakers normally take "corn" to apply to any type of grain, whereas American speakers take "corn" to apply exclusively to maize. Likewise, the term "defamation" has both a nontechnical meaning that laymen use in everyday, nonlegal contexts, and a strict legal meaning that lawyers use. Legal specialists are adept at keeping these two meanings separate, and avoiding treating the layman's use of "defamation" as obviously pertaining to the same topic as the legal one. Just how people manage delicate cognitive maneuvers like this is a difficult question. Their ability seems to turn in part on the fact that they are implicitly aware of there being two distinct representational traditions,

17. The automatic and largely subpersonal nature of linguistic coordination is a central explanandum for theories of linguistic communication. *See* Pinker (1994) for a vivid characterization. For a model of the process of coordinating mental files that does not rely on any metarepresentation, *see* Cumming (2013).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

and this higher-order awareness helping to maintain and regulate their deployments of the term in two different ways.

VII. CONTENT DETERMINATION

We have been using the term "topic" to refer to what is picked out by both concepts and linguistic meanings. Once we really get down to the details, however, there might be a need to distinguish between what concepts pick out and what meanings pick out. For one thing, there can be concepts without corresponding linguistic meanings, as was the case in our tram example above before the introduction of the name "Fabulous Fiona". There are also reasons to think that linguistic meanings are finer grained than concepts. For example, the words "moral" and "morality" may express the same underlying concept of a particular normative status, but the syntactic difference entail that the meanings of the two terms play different roles in a compositional semantics for a language.[18] In this section and the balance of this paper, let us focus on concepts and the topics they pick out, with the assumption that what we say about them should be applicable with little modifications for meanings and the topics they pick out. We will use the term "content" to refer to any topic that a concept picks out.

Contents are to concepts what semantic values are to linguistic meanings. In compositional semantics, semantic values of expressions are usually conceived as whatever items contribute to the truth-conditions of whole sentences. Analogously, we are thinking, contents are whatever items that contribute to the truth- or correctness-conditions of whole thoughts. Furthermore, the term "content", like "topic", is meant to be neutral between realist and anti-realist accounts of whatever is picked out by concepts. Traditional realists about a subject matter hold that what is picked out by the relevant concept is a stable object or property, while anti-realists deny this. Expressivists may take a content to be a distinct motivational profile; and fictionalists may take a content to be a metaphysically incoherent feature.[19] For these reasons, it would be question-begging to use the terms "reference" or "referent" across the board to talk about what is picked out by concepts.

Once again, sameness of concept is supposed to guarantee sameness of topic. But how exactly do concepts pick out topics as they are supposed to? Here, we delve into the second half of our overall theory of concepts. In the preceding sections, we have outlined our view of what it takes for two people, or one person at different times, to share a concept—i.e., what could be called our "theory of conceptual competence". In this section

18. *See* Pietroski (2010) and Glanzberg (2014) for articulations of this picture of the relation between concepts and linguistic meaning.

19. And contextualists and truth-relativists take a content to be a stable reference-fixing rule, which determines a reference on the basis of facts about the context of utterance or the context of assessment, respectively.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

and the next, we outline our take on how concepts determine contents. This portion of our view could be called our "theory of content determination".[20] In the case of representational concepts like SOCRATES or WATER, a content determination theory must explain what makes it the case that all deployments of these concepts single out a particular individual or a particular kind of substance as the content. An adequate determination theory must ensure that all possible deployments of the same concept will be assigned the same topic, irrespective of variations in the relevant thinker's mental states or environment. In the case of concepts with a stable reference, this means that the competence conditions must suffice for co-reference. A determination theory must provide general principles that explain why this is so.

What we are after here is a general account of how contents get assigned to concepts, so an adequate determination theory should apply across the board to all concepts and in a way that generates plausible verdicts for specific cases. The crucial point is that the theories of conceptual competence and content determination must fit together to ensure that stability of content for such concepts across all possible contexts in which they could be deployed.

The traditional approach to concepts favors the matching model not only for conceptual competence, but also for content determination. Or to put it another way, the matching model posits one entity—namely, a person's substantive understanding or conception of the subject matter of his thoughts—for explaining both his conceptual competence and the content determined by his thoughts. For example, a person counts as having BACHELOR thoughts in so far as he deems only eligible unmarried men as instances of the kind in question. And his thoughts, according to the model, pick out those who satisfy the criteria set out in that conception of bachelorhood. The entire picture is individualistic in the sense that sameness of concept and content determination are both secured by facts internal to the mind of the individual thinker at the time of the thought episode. On the traditional approach to concepts, facts external to the individual thinker's mind, including facts about the causal-historical relations among thought episodes, are considered irrelevant for both conceptual competence and content determination.[21]

20. In thus conceiving a theory of concepts as consisting of these two components, we are following Christopher Peacocke's (1992) seminal work on concepts.

21. More recently, neo-Fregean accounts of concepts have slightly modified this picture without thereby altering its fundamental commitment to individualism. Neo-Fregeans note that we often lack uniquely identifying conceptions that can single out the objects, kinds, or properties we keep track of in thought. One's understanding of elms, for instance, might not be sufficiently detailed to distinguish elms from beeches, to use Putnam's (1975) classic example. An individual thinker's internal conception of a topic therefore cannot be the only factor in determining the reference of an individual's thoughts and ensuring that every deployment of the same concept must co-refer. The neo-Fregean solution is to allow external factors to figure both in competence conditions and content determination. The reference of one's

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

In the preceding sections, we have argued against the matching model of conceptual competence, and its individualist commitments, in favor of the relational model. A commitment to a relational model of conceptual competence also precludes buying into the matching model's approach to content determination. In fact, the following three positions form an inconsistent triad:

(7) (a) the relational model of conceptual competence, which holds that people holding divergent substantive conceptions of a subject matter can deploy a single concept;

(b) the matching model of content determination, according to which an individual's substantive conception of a subject matter determines the contents of his concepts; and

(c) the thesis that sameness of concept guarantees sameness of content.

Since we are committed to both (7)(a) and (c), it follows that we must reject (7)(b); and an obvious alternative is the relational model of content determination.

Now, what would a relational theory of content determination look like? The simplest theory of such a sort would posit a simple relation of content inheritance. Take the case of discourse anaphora in our initial example of the conversation (1) about a modernist American female composer of the early twentieth century. The conversation here begins with A's use of the proper name "Ruth Crawford Seeger", which secures reference to a particular individual. Each subsequent participant in the conversation uses a pronoun that inherits its content from this original anchoring point. Saul Kripke's (1980) "causal chain" account of proper names seems to invoke such a quasi-anaphoric model of content inheritance: an initial baptism secures reference to an individual, and then the name is "passed from link to link" via co-referential intentions on the part of individuals when they originally acquire the name (p. 96). This account of content determination is anti-individualist in that it takes the content of a particular thought within a representational tradition to depend on historical facts about that

concept ELM, according to the proposed solution, could depend on one's disposition to interact with certain trees in one's physical environment, and/or on one's disposition to treat certain experts as authoritative. Despite the inclusion of such external factors, the approach ultimately sticks to the same individualistic approach to conceptual competence that the older Fregean model proffered: two thought episodes deploy the same concept iff each is independently associated with the same (externally individuated) set of dispositions. And the determination theory is similarly individualistic: a semantic content is assigned to a particular thought episode on the basis of the (externally individuated) dispositions governing that very thought episode in its empirical context. In effect, each thought episode must earn its reference off its own bat, and then sameness of concept is determined by a perfect match between these individual content-fixing achievements. For a more detailed discussion, *see* Schroeter (2012). In the text of this paper, we ignore this modified Fregean account.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

tradition, rather than on facts about the individual thinker's current conception of the topic.

This sort of content inheritance, however, cannot provide a general relational model of how sameness of content is secured within representational traditions. Although such a simple content-inheritance model predicts the intuitive content of many uses of proper names, the model does not fit all uses of names, and it certainly does not fit all expressions. For example, it notoriously struggles to explain shifts in the contents of proper names and natural kind terms over time (Evans 1973; Devitt 1981). Moreover, the idea that the contents of natural kind concepts or normative concepts, including legal concepts, are fixed by some original baptismal events is dubious. Different, more generally applicable principles for assigning contents are needed.

VIII. CONSTRUCTIVE INTERPRETATION

This is where, we suspect, something close to the Dworkinian conception of constructive interpretation could be helpful. Dworkin (1986, pp. 65–68; 2011, p. 131) distinguishes three interpretive steps:

(8) (a) identifying the relevant practice associated with a particular concept;
(b) identifying the practice's core purposes; and
(c) identifying the best satisfiers or realizers of those purposes.

The determination theory we sketch below respects the broad outlines of the interpretive process, as Dworkin conceives it, fleshing out crucial details and explaining how the approach constitutes a general account of how conceptual content is determined.

First, how should we demarcate the relevant "practice" associated with a particular concept? On a Fregean matching approach, a person participating in a shared "practice" associated with a concept would need to accept a specific conception of the relevant subject matter. On a simple content-inheritance model of the type we considered in the preceding section, a person would need to grasp quasi-syntactic rules for content inheritance. In devising a relational model that could be contrasted with both, we suggest that we begin by conceiving the relevant "practices" as demarcated by representational traditions. As we stressed earlier, representational traditions allow for open-ended variation in conceptions and patterns of use. What binds thought episodes into a single tradition is the fact that they are causally linked in the right way, not the fact that they resemble each other. This approach to demarcating shared "practices" places no strict constraints on the conceptions and patterns of use that can be associated with a particular tradition at different times or by different individuals.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

Our suggestion, then, is that the determination theory should seek to assign contents to a whole representational tradition considered as a corporate body. In effect, the determination theory treats an entire representational tradition as the basic *unit for interpretation*, the unit to which semantic content is assigned in the first place. Each individual thought episode earns its content in virtue of its being part of, and its particular location within, a larger, causally-historically extended representational tradition, rather than in virtue of the conception a thinker currently associates with that episode (in their current context). Taking whole traditions, rather than individual thought episodes, as the relevant units for interpretation means that content assignments can remain stable through significant variations in conceptions within a tradition. In effect, the determination theory pools all the information about the tradition as a whole in order to interpret each thought episode within that tradition.[22]

Now, what are the general interpretive principles that can provide a plausible content assignment for a shared representational tradition? Dworkin speaks of the "purposes" of the relevant social practices and specific interpretations that satisfy those purposes (2011, p. 131). Normally, we do not have specific purposes of practices in mind when we entertain thoughts within the practice, or express them in language. Instead, our attention is simply focused on the topics picked out. We do, however, appeal to something like purposes when we step back and try to clarify what exactly we are thinking and talking about. This is especially likely to happen, for example, when we encounter hard cases that challenge our previously held ordinary understanding of a topic, when we are engaged in systematic inquiry into the nature of a familiar topic, or when we find that we disagree about the proper definition of a shared concept. We suggest that contents can be assigned to concepts—e.g., FREE WILL, LEGAL, DEFAMATION, etc.—by next attributing points or purposes to the relevant representational traditions. More specifically, the goal is to determine the most important shared *representational interests* that are served by these representational traditions. Why has it been important to the relevant community to participate in these representational traditions? And what justifies the community's continued engagement in and maintenance of the tradition?

Perhaps an example will clarify what we have in mind with "representational interest". Consider the philosophical debate about the nature of

22. A qualification is in order here. Strictly speaking, the representational tradition is the default unit for semantic interpretation. That is, the determination theory has a strong initial bias in favor assigning a univocal interpretation to the tradition as a whole. However, this initial bias can be defeated. Some representational traditions are ambiguous and some traditions shift their contents over time. We shall ignore such complications in this paper. For discussion of disambiguating interpretations, *see* Schroeter (2008); Schroeter & Schroeter (2015). The key point is that disambiguating a representational tradition is not a matter of matching internal conceptions associated with individual thought episodes within that tradition. Even within the process of disambiguation, the determination theory privileges continuous representational traditions over individual thought episodes considered independently.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

race. What are we attributing when we characterize people as black, yellow, or white?[23] In answering this question, some theorists treat *biological explanatory interests* as the most important. Both children and adults tend to treat racial categories as base-level biological kinds, positing internal properties passed from one generation to the next that ground a wide range of predictions and explanations of manifest properties. But because the biological theory does not vindicate the relevant sorts of essentialist explanations, such theorists often argue that racial categories are empty.[24] In contrast, social constructivists about race focus on the role that racial categories play in social domination. For example, classifying someone as "black" tends to situate that person in a social hierarchy and activates a suite of negative stereotypes and discriminatory norms. This *interest in locating individuals within a social hierarchy* grounds very different interpretations of the nature of racial categories. Being black, yellow, or white does not depend on biological characteristics. Instead, it is a marker of social status tied to certain manifest properties.[25] Thus in debates over the nature of racial categories, disputants seek to identify the most important representational interests at stake in our shared representational traditions, which they then use as a constraint on adequate definitions of the categories in question.

Finally, the third step is to determine the optimal way, if any, to satisfy those representational interests. Adjudications among different interpretations of some concept can therefore happen at either the second or third step. The contest among different interpretations could be settled by attributing different points or purposes, or by different proposals to serve whatever points or purposes are selected.

In sum, as we see it, there are three key aspects of our relational approach to content determination, which neatly map onto Dworkin's account of constructive interpretation that we outlined in (7) above:

(9) (a) *Unit of interpretation*: The unit for semantic interpretation is the whole representational tradition.

23. Our excellent copyeditor, Anna Skiba-Crafts, alerted us to the fact that many consider "yellow" a slur rather than a neutral descriptor. With some trepidation, we decided to retain its use for several reasons. First, there is no obvious good alternative. Second, our small informal poll of younger and more "woke" colleagues indicates that people have different intuitions about the term's tone, especially in contexts like ours. Third, even among those who believe that it is a slur, there appears to be a desire to reclaim it as a neutral descriptor. On this last point, *see* Chow (2018). Thanks to Daniel Wodak for this reference, as well as an informative e-mail exchange about the issue.

24. Appiah (1996) and Zack (2002) argue for eliminativism about standard racial concepts for these sorts of reasons. In contrast, Kitcher (2007) argues for a modest biological account of race that defines racial categories in terms of broad genealogical lineages that ground superficial biological kinds. Such categories would in effect be analogous to different breeds in dogs.

25. *See* Haslanger (2000) and Mallon (2006) for two versions of such social constructivist theories of race.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

(b) *Representational interests*: Interpretation seeks to determine the most important shared representational interests at stake in justifying the continuation of the representational tradition.

(c) *Analysis*: Interpretation seeks to define which feature or item, if any, best satisfies those interests.

We suspect that discussion of some examples would facilitate an understanding of what we are proposing here, and we will now work through a couple of them.

IX. PIRACY

The spate of piracy off the coast of Somalia in the last two decades has instigated renewal of some old legal debates about the definition of "piracy" in American federal courts.[26] Historically, there have been two legal categories of piracy—municipal piracy and general piracy. Municipal piracy covers any acts deemed by the legislature of some state as acts of piracy. A state may prosecute individuals for such municipal piracy only when some of the traditional bases of jurisdiction are satisfied—e.g., when their conducts occur within the territory of the relevant state, when the conducts occurring outside the territory are by the nationals of the state. In contrast, general piracy can be prosecuted by any state, irrespective of any jurisdictional nexus. But unlike municipal piracy, general piracy is restricted to those offenses that are classified as piracy by consensuses existing in the international community.

The U.S. Constitution provides that "[t]he Congress shall have Power . . . [t]o define and punish Piracies and Felonies committed on the high Seas, and Offences against the Law of Nations".[27] The recent federal cases in the Fourth Circuit concern Section 1651 of Title 18 of the U.S. Code, according to which "[w]hoever, on the high seas, commits the crime of piracy as defined by the law of nations, and is afterwards brought into or found in the United States, shall be imprisoned for life". Congress has thus incorporated within the U.S. Code the international legal definition of general piracy. The recent controversy in the Fourth Circuit had to do with whether the definition of "piracy" in Section 1651 evolves with developments in the customary international law, or whether it has been static since the seminal 1820 U.S. Supreme Court decision in *United States v. Smith*.[28] The Court in *Smith*, examining Section 5 of the Act of 1819, which was the predecessor of 18 U.S. Code §1651, stated that by incorporating the definition of piracy under the law of nations, Congress had defined "piracy" as clearly and determinately as if it itself had enumerated the elements of the offense.

26. Thanks to Alex Sarch for pointing us to this example.
27. U.S. Const. art. I, §8, cl. 10.
28. 18 U.S. (5 Wheat.) 153 (1820).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

And according to that definition, the Court opined, piracy was "robbery, or forcible depredations upon the sea, *animo furandi*".[29]

The relevant representational tradition here consists not merely of the American domestic legal tradition that has dealt with piracy, but also of the wider public international legal tradition within which that domestic legal tradition is located. Section 1651 explicitly incorporates the customary international law governing general piracy. The obviously central representational interest in thinking and talking about general piracy in this tradition is to designate various conducts that occur in high seas as crimes that any state in the world can prosecute and that any courts in the world would have the jurisdiction to try. But there is another significant representational interest that has motivated district courts in the Eastern District of Virginia to take different approaches to defining "piracy". This is not an interest in sustained use of the concept PIRACY in particular, but instead an interest in sustained use of a panoply of more general concepts we use in thinking and talking about criminal law, and even more generally about the politics of ordered liberties. This is the interest in furnishing potential criminal defendants with sufficiently clear warning as to what kinds of behavior amount to crimes, and in American constitutional law this due process interest is enshrined in the Fifth Amendment of the U.S. Constitution.

The due process interest appears to have played a decisive role in the district court's decision in *United States v. Said*.[30] That case involved Somali pirates who mistook a U.S. naval vessel in the Gulf of Aden for a merchant ship and began to fire on it. The naval vessel returned fire, and soon captured the Somali pirates and brought them to Virginia to be prosecuted. The district court decided that the defendants could not be prosecuted under Section 1651 because their actions did not amount to robbery or seizure of a ship as required by *Smith*. In effect, the *Said* court froze the definition of "piracy" in Section 1651 to the one formulated by the *Smith* Court in 1820, and it did so out of due process concerns. It took note of the evolution of the definition in customary international law since 1820, but came to the conclusion that there was insufficiently clear consensus in contemporary customary international law to provide sufficiently clear warning to potential defendants. A mere two months later, in *United States v. Hasan*,[31] the same district court in the Eastern District of Virginia, under a different judge, ruled that Somali pirates whose conduct matched those of the *Said* case almost exactly could be prosecuted under Section 1651. The *Hasan* court took into consideration the due process interest in furnishing fair warning, but came to the conclusion that there is

29. *Id.* at 161. The Latin phrase "animo furandi" means "with the intent to steal".
30. 757 F. Supp. 2d 554 (2010).
31. 747 F. Supp. 2d 599 (2010).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

sufficiently clear consensus in contemporary customary international law to satisfy that interest. In effect, the court ruled that the due process considerations did not freeze the definition of "piracy" in Section 1651. In so ruling the court appealed to, among others, (i) the decision of the Privy Council of England in *In re Piracy Jure Gentium*,[32] (ii) the Kenyan High Court's decision in *Ahmed v. Republic*,[33] (iii) Article 15 of the Geneva Convention on the High Seas,[34] and (iv) Article 101 of the United Nations Convention on the Law of the Sea (UNCLOS).[35] All of these sources indicate that general piracy is not limited to robbery, but can include armed assaults on ships without actual takings of any property. The *Hasan* court deemed these indications sufficiently clear to satisfy any due process concerns, and thus ruled that defendants who attack ships can be prosecuted under Section 1651. This decision was later affirmed by the Fourth Circuit in *United States v. Dire*.[36]

The dispute between the *Said* and the *Hasan* courts was not driven by their commitments to different and competing representational interests. Both courts in determining the content of PIRACY or GENERAL PIRACY took into account an interest in prohibiting certain kinds of behavior at high seas and enabling states everywhere to prosecute individuals who take part in such behavior, and also the due process interest in furnishing potential defendants with sufficiently clear warning. But it appears that the courts differed in their assignments of weights to the two interests. The *Said* court assigned so much weight to the due process interest as to require a very high degree of consensus in contemporary customary international legal conception of piracy, whereas the *Hasan* court seems to have required a lower threshold for such a consensus to meet the due process requirement. The different assignments of weights to the two interests clearly yielded different contents for the concept PIRACY or GENERAL PIRACY. The *Said* court's approach limited piracy to robberies at sea and seizures of ships, whereas the *Hasan* court's approached yielded a content that includes a greater varieties of armed assaults at high seas. Given the arbitrariness of the distinction between robberies and seizures of ships on the one hand and other sorts of seaborne armed assaults on the other for the purposes of satisfying the first interest, and also the largely fictional nature of any fair warnings that potential pirates would receive, we ourselves find it hard to take issue with the *Hasan* court's approach. And on appeal, the Fourth Circuit appears to have largely agreed.

32. [1934] A.C. 586 (1934).
33. Crim. App. Nos. 198, 199, 201, 203, 204, 205, 206 & 207 of 2008 (H.C.K. May 12, 2009).
34. Apr. 29, 1958, 13 U.S.T. 2312, 450 U.N.T.S. 397.
35. Dec. 10, 1982, 1833 U.N.T.S. 397.
36. 680 F.3d 446 (4th Cir. 2012). The district court's decision in *Said* was later overturned accordingly. United States v. Said, 680 F.3d 374 (4th Cir. 2012).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

X. CULPABILITY

Our second example comes not so much from actual legal disputes in courts, but instead from some scholarly discussions about criminal law. It is a core commitment of civilized legal systems that an individual should not normally be convicted of a criminal offense unless he is *culpable* or *blameworthy* with respect to that offense.[37] Criminal law has traditionally dealt with this issue by findings of mens rea. But exactly how such findings constitute or indicate culpability is a difficult problem that is likely to have far-reaching implications in the actual criminal legal practice.[38]

The issue of culpability is a very deep and large one that has generated a sizeable literature over the years. There is no good way to take all of it in in a short discussion that is meant only to illustrate how the content of the concept CULPABILITY is likely to be determined. Dispensing with any proper scene-setting, we shall here take for granted A.P. Simester's (2013, pp. 179–180) recent conjecture that the point or purpose of findings of culpability is to trace our negative evaluation of some act back to the agent. As Simester explains:

> [M]oral evaluations of an act are not the same kind of evaluation as blaming judgments. Their objects of assessment are crucially different. When we blame, we disapprove not of the act by itself, but *of the defendant for doing that act*. . . . The burden for ascribing culpability lies in justifying that evaluative link between act and defendant. . . . To blame someone is to make a moral assessment of that person *in respect* of their action. The challenge for culpability is . . . a challenge about how to generate judgments that integrate evaluations of the actor and her act. (p. 180)

Following Simester, we can here take the central representational interest in sustaining thought and talk about culpability or blameworthiness as that of integrating evaluations of agents with their acts in criminal legal settings.

Over the years, a number of criminal law theorists and philosophers have devised theories of culpability. And we can see these theories as proposals for the best satisfiers or realizers of the central representational interest of integrating evaluations of agents with their acts in criminal legal settings. According to an influential line of reasoning (*e.g.*, Moore 1990; Moore & Hurd 2011), a necessary condition for culpability is that the defendant made a choice of carrying out a wrongful act (or of omitting to carry out the right act). As Simester (2013, p. 190; *cf.* 2000) points out, such a choice theory of culpability ill fits culpability for inadvertent negligence. As he elaborates:

37. We thank Mark Dsouza for pointing us to this example.
38. *See generally* Chan & Simester (2011).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

> Chosen wrongdoing reflects badly on the actor because it shows that he prefers bad reasons to good ones. He would rather stab his opponent than walk away, and so on. By contrast, the complaint when D is negligent is, in a sense, that she fails to act for the right reasons. But this deficiency, too, can reflect badly upon her *qua* rational moral agent. This is because the capacity for rational deliberation does not just involve capacities to assess moral reasons (and act for undefeated ones); it also involves capacities to recognize them. (2013, p. 191)

This observation indicates that the connection between inadvertent negligence and culpability is likely to bypass any choice on the defendant's part. Choice theories, or at least simple versions, appear to fail to delineate the best satisfiers or realizers of the aforementioned central representational interest of using the concept CULPABILITY. Simester (pp. 192–193) proposes that we should opt for a "disintegrated theory", according to which culpability may be grounded either in a choice to act on bad reasons, or in certain epistemic failures—e.g., a failure to take into account or attend to certain risks—on the part of the defendant. Some epistemic failures, he reasons, displays certain flaws of character that warrant findings of culpability.

While we are hardly in a position to adjudicate this debate about the nature of culpability or blameworthiness, we take the dialectic we have thus sketched to illustrate the kind of interpretive processes that go into content determination.

XI. COMPARISON WITH DWORKIN

The relational model we have so far sketched underwrites some elements that could be deemed Dworkin's core commitments about concepts. Dworkin's view of concepts was intended to explain a puzzle about genuine agreements and disagreements in domains like law, politics, and morality, where individuals often have wildly divergent understanding of the natures of the topics picked out (1986, ch. 1; 2011, pp. 6–7, 157–158). If individuals do not share the same criteria for deciding which things count as "legal", "democratic", or "morally right", the thinking went, how can their uses of those terms express the same concepts and thus ground strict logical disagreements about the very same topic? We have shown how a relational model of concept identity can offer a solution. Unlike traditional accounts of concepts, accounts that adhere to the relational model of concepts are compatible with variation, even very wide variation, in any aspect of understanding. The relational model thus solves the puzzle that motivated Dworkin's account.

Moreover, the relational model captures the structure of Dworkin's account of interpretive concepts. On the relational model we have sketched, competence with a shared concept requires participation in a

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

shared representational practice within a linguistic community. And the semantic content of a concept is determined by way of a rationalizing interpretation. Such interpretation involves (i) treating the practice as the default unit for interpretation, (ii) identifying the representational interests served by that practice, and (iii) determining which content assignment, if any, would best meet those interests. The outlines of such a conception of interpretation capture the spirit and even some of the details of Dworkin's programmatic discussion of interpretive concepts.

The relational model also underwrites Stravopoulos's (1996) important observation that Dworkin's rejection of "criterial concepts" aligns with the rejection of individualism, and the search for an anti-individualist approach to concepts. The relational model is clearly anti-individualist as it characterizes concept identity as depending essentially on belonging to the same representational tradition. According to it, two token elements of thought express the same concept only if they are causally-historically connected by chains of apparent guaranteed sameness relations. It follows that it is not the case that what happens inside an individual's cognitive economy determines whether that individual uses the word "legal" to express the same concept as those of others. In addition to any purely internal facts about that individual's substantive understanding of the topic, the token element of thought he expresses must be causally-historically related to that of others to count as expressing the same concept. It follows that competence with a given concept is not something that individuals can achieve on their own, independently of a connection to others or their past selves via a representational practice.

There are, to be sure, also some important differences between the proposed relational model of concepts and Dworkin's account of concepts. First, Dworkin distinguishes "interpretive concepts" from both "criterial" and "natural kind concepts". He is thereby committed to there being multiple mechanisms of concept identity. Dworkin's programmatic account of this contrast leaves important questions unanswered. For one, what determines whether a particular token element of thought expresses a criterial, causal, or interpretive concept? Also, how do the different types of concept fit within an overall theory of concept identity? In contrast, the relational model we outlined above is a general theory of concept identity. It does not cover only normatively laden topics such as law, politics, and morality. Instead, it is meant to cover all concepts, including concepts that pick out natural kinds and artifact kinds, concepts that attribute observable or theoretical properties, and even concepts that express logical relations. According to the relational model, the key requirement for sharing *any* concept is not the grasp of a precise content-fixing criterion, but participation in a representational tradition. Rationalizing interpretation of the point of a representational practice, and assessing candidates for serving that purpose, determines the content of *any* concept. Thus, an interpretation must determine the dominant representational interests at stake in justifying a

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

particular representational practice—e.g., whether the point of a particular classificatory practice is predominantly normative, causal-explanatory, superficially descriptive, expressive, or what have you. Some of our representational practices—such as our practice with GOLD—are best interpreted as aimed at keeping track of information about fundamental causal-explanatory kinds with which we causally interact. Others are best interpreted as aimed at keeping track of socially important kinds (possibly, CAUCASIAN), or kinds that play normative roles (e.g., MORALLY RIGHT), or kinds that capture superficial observable features (e.g., RED). Thus, the distinction that Dworkin marks between concepts with noninterpretive semantics can be drawn instead within the relational model as derived from a general interpretivist theory of content determination. The relational model generalizes the interpretive approach in such a way as to provide a coherent overall framework for thinking about concepts that pick out a variety of different types of reference, as well as nonrepresentational contents. This generality strikes us as an important advantage of the relational model over Dworkin's interpretivism.

Another important difference between the proposed relational model of concepts and Dworkin's has to do with normative realism. There has been a long streak of ambivalence about normative realism in Dworkin's work (*e.g.*, 1996 & 2011), and his interpretivist conception of normative concepts could be seen as a product of that ambivalence. At one point in *Justice for Hedgehogs*, Dworkin says that in the face of evaluative conflicts, we do and should resort to repeated reinterpretations or "rearrangements" of key evaluative concepts. He continues:

> What other story might one tell? Consider this one: "Moral conflict is real, and any theory that denies this is false to moral reality. Once we understand the nature of kindness and of honesty, we see that, in cases like this one, they just *do* conflict. That conflict is not an illusion produced by incomplete moral interpretation; it is a matter of plain fact." But what in the world could that supposed plain fact consist in? Kindness and honesty cannot *just* have one content or another, because moral claims cannot be barely true. I repeat tediously: no moral particles fix what these virtues just are. Nor do the concepts have a precise and conflicting content just in virtue of linguistic practice. Moral concepts are . . . interpretive concepts: their correct use is a matter of interpretation, and people who use them disagree about what the best interpretation is. (2011, p. 120)

Of course, what Dworkin calls "plain fact" and "bare truth" here could be taken in a variety of ways. Still, what clearly come across are Dworkin's misgivings about straightforward realism about what moral or normative claims are about. His interpretivist conception of normative concepts appears to be designed to enable him to steer clear of both such realism and some crude conventionalism ("just in virtue of linguistic practice"). But the

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

prospect of any such navigational feat depends crucially on the assumption that interpretivism is characteristic only of normative concepts, while some other approach is appropriate for concepts about paradigmatically realist topics such as natural kinds or artifacts. The distinction between interpretivist concepts and other types of concept is supposed to explain why there are "plain facts" about such matters as gold and sofas, but not about defamation and elegance.

It should be clear from our foregoing discussion that the relational model of concepts provides no support for such avoidance of normative realism. As a general theory of conceptual competence and content determination, the relational model itself is neutral about the metaphysical character of the different topics picked out by concepts. In particular, reliance on interpretation to assign contents does not entail anything about the metaphysical character of the contents assigned. But this neutrality does not imply or motivate any avoidance of any metaphysical commitments. The best rationalizing interpretation of the thoughts we express with "gold", for instance, is a perfectly objective chemical kind, Au. And the best rationalizing interpretation of our tradition with "Aristotle" is a perfectly objective historical human being born in Stagira in the fourth century BC. It is important here to be very clear about the distinction between two types of theory:

(10) (a) *A content determination theory*: a theory of how semantic contents are assigned to token elements of thought on the basis of empirical facts about that token—including facts about its role in the individual's cognitive economy and facts about its relations to its historical, social, and physical contexts.

(b) *A metaphysical theory*: a theory about the metaphysical nature of the features that concepts represent—i.e., the metaphysical nature of the concept's reference, or putative reference.

To claim that the relational model's reliance on an interpretivist determination theory is sufficient to rule out normative realism is to conflate a determination theory with a metaphysical theory. A determination theory is inevitably mind-dependent, since it assigns contents as a function of facts about token elements of thought. Which feature, if any, is picked out as the content of a token element of thought depends at least in part on the thinker's understanding of the topic. But it does not follow that the intrinsic nature of the features picked out by the thoughts is mind-dependent.

A related point concerns Dworkin's views about the nature of law. When Dworkin first introduced his interpretivism, he accused both legal positivists and natural law theorists of faulty semantics (1986, esp. ch. 1). Subsequently, however, Dworkin and many of his readers have given the impression that there is a rather short distance between interpretivism

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

and natural law theory. This last move would be unwarranted, however, if we were to develop interpretivism along the trajectory we are projecting. An argument for natural law theory must appeal to the details of the shared representational practice expressed by the concept LAW. For any concept, an interpretivist determination theory is committed to taking into account the entire set of attitudes, dispositions, social practices, and environmental feedback loops that make up the historically and socially extended representational practice. To determine the right content assignment for LAW and other legal concepts, we should seek to identify *the point* of the legal representational practice—i.e., elucidating the dominant representational interests at stake. Answering this question is bound to be both difficult and controversial. Roughly, legal positivists highlight our interest in keeping track of objective social commitments, whereas natural law theorists highlight the role of moral justifiability of our legal practices. Which of these interpretations (or more accurately, any member of the two families of interpretation) better captures the interests at stake in our legal practices cannot be adjudicated at the level of semantic theorizing—it depends crucially on a detailed analysis of our shared legal institutions and practices, on an explanatory analysis of how they function, and on a normative assessment about which aspects of those practices are most important to us as participants. We leave open the possibility that the point of the legal representational tradition will have moral implications that motivate a commitment to a natural law conception of the nature of law.[39] But any such implications will have to be shown by a sustained discussion and arguments, and we do not see an easy or short way to them from the theory of concepts we have outlined here.

Finally, Dworkin's ambivalence about normative realism is part and parcel of his general skepticism about metaethics or meta-normative theorizing more generally. His arguments against meta-normative theorizing over the years have been quite complex, multi-faceted, and at times bewildering. One key assumption driving the arguments has been the view that any meta-normative theorizing invariably has the "Archimedean" ambition of vindicating or debunking normative thinking wholesale. Another driving assumption has been the thesis of the "autonomy" or "independence" of normative thinking. Inspired by Hume's *is-ought* gap, Dworkin (*e.g.*, 2011, chs. 3–5) has repeatedly argued over the years that our justification for affirming or denying first-order normative claims must be based on first-order normative premises. Insofar as meta-normative theorizing, as he conceives it, seeks to either vindicate or debunk first-order normative thinking based on nonnormative considerations alone, Dworkin opines, it is idle and bogus. Thus his insistence that any genuine normative skepticism must be an "internal skepticism" (1986, pp. 76–85; 1996; 2011, chs. 3–5). While

39. For some recent discussions that suggest some such implications, *see, e.g.*, Copp (2019); Railton (2019).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

eschewing the "Archimedean epistemology", which he suspects any meta-normative theorizing as invariably involving, Dworkin has forcefully advocated and practiced what he calls "integrated epistemology" within normative domains of thought (2011, pp. 82–83). "Integrated epistemology" is none other than the epistemology of reflective equilibrium. But given Dworkin's assumption of normative autonomy, it is *narrow* reflective equilibrium. Dworkin has advocated setting aside nonnormative considerations, including meta-normative considerations, in pursuing a set of integrated commitments in first-order normative theorizing.

Many moral and legal philosophers would be sympathetic to Dworkin's defense of the authority of first-order normative thinking and first-order legal argument.[40] The relational model of concepts provides partial support for this strand of Dworkin's "anti-metaethical" position. According to the interpretivist theory of content determination, the key to specifying the content of our legal and normative concepts—and indeed to specifying the content of any concept—is to engage with the details of our representational practices with those concepts. Any attempt to undermine those practices by appealing to metaphysics or empirical science will fail unless it can explain why the metaphysical or scientific phenomena are crucial to the representational interests at stake in our ongoing representational practice. For instance, why should we care if our moral judgments do not track a naturalistic property? If the meta-normative skeptic cannot explain why those who use normative concepts should care about the allegedly debunking facts, it is hard to see how the allegedly debunking facts could really be central to the representational interests that justify our representational practices with moral concepts. Thus, the type of determination theory we have sketched, which invokes rationalizing interpretation of the point of a practice, will ultimately be answerable to the ideal reflective judgments of participants in that practice. An interpretation cannot be correct if it does not make sense from the perspective of concept users themselves. In this sense, the relational model of concepts we have sketched lends support to Dworkin's view that first-order normative thinking takes precedence over generic metaphysical, semantic, or scientific thinking in adjudicating skeptical challenges.

The relational model, however, does not support the stronger Dworkinian claim that first-order normative thinking is entirely autonomous from, and hence is not answerable to, general meta-normative theorizing. Indeed, this claim seems unwarranted by Dworkin's own lights. Dworkin's account of interpretive concepts is intended to show how logical disagreement about legal and normative questions is possible, despite open-ended variation in individuals' substantive understanding of those topics. Dworkin clearly feels the force of skeptical worries about first-order normative disagreement, even when such skeptical worries are based on general

40. *See, e.g.*, Nagel (1997, esp. ch. 6); Scanlon (1998, ch. 1). *See also* Rawls (1975).

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

semantic views about conceptual competence and content determination. In sketching his alternative interpretivist view of concepts, he sought to assuage these skeptical worries and vindicate our first-order normative practices. In other words, Dworkin's meta-level theory of interpretive concepts is motivated by his desire to vindicate ordinary first-order normative thinking. But what if Dworkin's programmatic account of interpretive concepts proves inadequate to that task? If there is no way of filling in his interpretivism in such a way as to vindicate genuine logical disagreement, then we would need to reassess the status of deep normative and legal disputes. Perhaps we should say that the discussants involved in such disputes are using the same words to express distinct concepts, and thus simply talking past each other. Or perhaps we should say that the discussants are right in assuming that they share a concept, but that their shared concept is empty because nothing can satisfy their conflicting opinions. In either case, meta-normative considerations will have significant consequences for first-order normative claims. So normative autonomy does not seem consistent with original motivations of Dworkin's own account of concepts. In the absence of a detailed theory of concepts, Dworkin remains vulnerable to the critique that he has failed to deliver the "honest toil" needed to vindicate his anti-skeptical view about normative and legal disagreements. In this paper, unlike Dworkin, we have taken up the task of vindicating the anti-skeptical view about normative and legal disagreements. We have sought to fill in a more general theory of concepts that would allow discussants to logically agree and disagree with each other about a topic despite differences, even quite radical differences, in their substantive understandings of the topic.

XII. CONCLUSION

It could be argued that the subfield of philosophy that deals with the nature of law, the subfield that often goes by the label "general jurisprudence", has far too long proceeded and developed in relative isolation, without the benefits of contributions and disciplining provided by the rest of philosophy and neighboring empirical disciplines. In particular, the suite of theoretical positions that Dworkin has developed over the years in his many seminal writings has gained traction and favor among many legal philosophers and legal theorists more generally without getting the kind of vetting, disciplining, and developments that philosophical positions often get. We share many legal philosophers' enthusiasm for the basic outline and some key components of Dworkin's positions. But we believe that the outline and the components can be arranged together in a way that yields a product that is quite different from what Dworkin and his many followers favor. What holds our arrangement together is an interpretivist metasemantics of the sort that Dworkin did not envision, but one that we believe that he should have adopted. We are pretty sure that there are other theoretical

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

options that can be, and have been, developed to explain the phenomenon of fundamental legal disagreements. But the one delineated in this paper, we believe, is worth scrutinizing and developing further as we consider other desiderata of success in explaining the nature of law.

REFERENCES

APPIAH, K. ANTHONY. 1996. *Race, Culture, Identity: Misunderstood Connections*. In COLOR CONSCIOUS, ed. K. ANTHONY APPIAH & AMY GUTMAN, 30–105. Princeton: Princeton University Press.

BURGE, TYLER. 1998. *Memory and Self-Knowledge*. In EXTERNALISM AND SELF-KNOWLEDGE, ed. PETER LUDLOW & NORAH MARTIN, 351–371. Palo Alto: CSLI Publications.

CHALMERS, DAVID. 2006. *The Foundations of Two-Dimensional Semantics*. In TWO-DIMENSIONAL SEMANTICS: FOUNDATIONS AND APPLICATIONS, ed. MANUEL GARCIA-CARPINTERO & JOSEP MACIA, 55–140. Oxford: Oxford University Press.

CHAN, WINNIE & A.P. SIMESTER. 2011. *Four Functions of Mens Rea*. CAMBRIDGE LAW JOURNAL 70: 381–396.

CHOW, KAT. 2018. If We Called Ourselves Yellow. https://www.npr.org/sections/codeswitch/2018/09/27/647989652/if-we-called-ourselves-yellow?t=1595628472058.

COPP, DAVID. 2019. *Legal Teleology: A Naturalist Account of the Normativity of Law*. In DIMENSIONS OF NORMATIVITY: NEW ESSAYS ON METAETHICS AND JURISPRUDENCE, ed. DAVID PLUNKETT, SCOTT SHAPIRO & KEVIN TOH, 45–64. Oxford: Oxford University Press.

CUMMING, SAMUEL. 2013. *From Coordination to Content*. PHILOSOPHERS' IMPRINT 13: 1–17.

DEVITT, MICHAEL. 1981. DESIGNATION. New York: Columbia University Press.

DICKIE, IMOGEN. 2010. *We Are Acquainted with Ordinary Things*. In NEW ESSAYS ON SINGULAR THOUGHT, ed. ROBIN JESHION, 213–245. Oxford: Oxford University Press.

DOWELL, J.L. 2015. *The Metaethical Insignificance of Moral Twin Earth*. OXFORD STUDIES IN METAETHICS 11, ed. RUSS SHAFER-LANDAU, 1–27. Oxford: Oxford University Press.

DWORKIN, RONALD. 1986. LAW'S EMPIRE. Cambridge, MA: Harvard University Press.

DWORKIN, RONALD. 1996. *Objectivity and Truth: You'd Better Believe It*. PHILOSOPHY AND PUBLIC AFFAIRS 25: 87–139.

DWORKIN, RONALD. 2011. JUSTICE FOR HEDGEHOGS. Cambridge, MA: Harvard University Press.

EVANS, GARETH. 1973. *The Causal Theory of Names*. PROCEEDINGS OF THE ARISTOTELIAN SOCIETY supp. 47: 187–208.

EVANS, GARETH. 1982. THE VARIETIES OF REFERENCE. New York: Clarendon Press.

FOOT, PHILLIPPA. 1959. *Moral Beliefs*. PROCEEDINGS OF THE ARISTOTELIAN SOCIETY 59: 83–104.

FORBES, GRAEME. 1989. *Cognitive Architecture and the Semantics of Belief*. MIDWEST STUDIES IN PHILOSOPHY 14: 84–100.

FREGE, GOTTLOB. 1892. *On* Sinn *and* Bedeutung. In THE FREGE READER, ed. MICHAEL BEANY, 151–171. Oxford: Blackwell, 1997.

GIBBARD, ALLAN. 1992. *Moral Concepts: Substance and Sentiment*. PHILOSOPHICAL PERSPECTIVES 6 (Ethics): 199–221.

GIBBARD, ALLAN. 2003. THINKING HOW TO LIVE. Cambridge, MA: Harvard University Press.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

Glanzberg, Michael. 2014. *Explanation and Partiality in Semantic Theory.* In Metasemantics: New Essays on the Foundations of Meaning, ed. Alexis Burgess & Brad Sherman, 259–292. Oxford: Oxford University Press.

Hare, R.M. 1952. The Language of Morals. Oxford: Clarendon Press.

Hare, R.M. 1986. *A* Reductio ad Absurdum *of Descriptivism.* In Philosophy in Britain Today, ed. Stuart Shanker. London: Croom Helm.

Hart, H.L.A. 1961/1994. The Concept of Law, 2nd ed. Oxford: Clarendon Press.

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

Heim, Irene. 1982. The Semantics of Definite and Indefinite Noun Phrases. Ph.D. dissertation in Linguistics, University of Massechussetts, Amherst.

Horgan, Terence & Mark Timmons. 1992. *Troubles on Moral Twin Earth: Moral Queerness Revived.* Synthèse 92: 221–260.

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

Jeshion, Robin. 2010. *Singular Thought: Acquaintance, Semantic Instrumentalism, and Cognitivism.* In New Essays on Singular Thought, ed. Robin Jeshion, 105–140. Oxford: Oxford University Press.

Kamp, Hans. 1981. *A Theory of Truth and Semantic Representation.* In Formal Methods in the Study of Language, ed. J. A. G. Groenendijk, T. M. V. Janssen & M. B. J. Stokhof, 277–322. Amsterdam: Mathematical Centre.

King, Jeffrey C. & Karen S. Lewis. 2018. *Anaphora.* In Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta. https://plato.stanford.edu/archives/fall2018/entries/anaphora/.

Kitcher, Philip. 2007. *Does 'Race' Have a Future?* Philosophy and Public Affairs 35: 293–317.

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

Lawlor, Krista. 2001. New Thoughts About Old Things: Cognitive Policies as the Ground of Singular Concepts. New York: Garland Publishing.

Lewis, David. 1970. *How to Define Theoretical Terms.* Journal of Philosophy 67: 427–446.

Mallon, Ron. 2006. *Race: Normative, Not Metaphysical or Semantic.* Ethics 116: 525–551.

Millikan, Ruth Garrett. 2000. On Clear and Confused Ideas. Cambridge: Cambridge University Press.

Millikan, Ruth Garrett. 2017. Beyond Concepts: Unicepts, Language, and Natural Information. Oxford: Oxford University Press.

Moore, Michael. 1990. *Choice, Character and Excuse.* Social Philosophy and Policy 7: 29–58.

Moore, Michael & Heidi Hurd. 2011. *Punishing the Awkward, the Stupid, the Weak, and the Selfish: The Culpability of Negligence.* Criminal Law and Philosophy 5: 147–198.

Nagel, Thomas. 1997. The Last Word. New York: Oxford University Press.

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

Perry, John. 1980. *A Problem About Continued Belief.* Pacific Philosophical Quarterly 61: 317–332.

Perry, John. 2001. Reference and Reflexivity. Palo Alto: CSLI Publications.

Pietroski, Paul. 2010. *Concepts, Meanings, and Truth: First Nature, Second Nature, and Hard Work.* Mind and Language 25: 247–278.

Pinker, Stephen. 1994. The Language Instinct: How the Mind Creates Language. New York: William Morrow.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

PLUNKETT, DAVID & TIM SUNDELL. 2013a. *Disagreement and the Semantics of Normative and Evaluative Terms.* Philosophers' IMPRINT 13: 1–37.

PLUNKETT, DAVID & TIM SUNDELL. 2013b. *Dworkin's interpretivism and the Pragmatics of Legal Disputes.* LEGAL THEORY 19: 242–281.

PUTNAM, HILARY. 1975. *The Meaning of 'Meaning'.* MINNESOTA STUDIES IN THE PHILOSOPHY OF SCIENCE 7: 131–193.

RAILTON, PETER. 2019. *"We'll see you in court!": The Rule of Law as an Explanatory and Normative Kind.* In DIMENSIONS OF NORMATIVITY: NEW ESSAYS ON METAETHICS AND JURISPRUDENCE, ed. DAVID PLUNKETT, SCOTT SHAPIRO & KEVIN TOH, 1–22. Oxford: Oxford University Press.

RAWLS, JOHN. 1975. *The Independence of Moral Theory.* PROCEEDINGS AND ADDRESSES OF THE AMERICAN PHILOSOPHICAL ASSOCIATION 48: 5–22.

RECANATI, FRANÇOIS. 1993. DIRECT REFERENCE: FROM LANGUAGE TO THOUGHT. Oxford: Blackwell.

RECANATI, FRANÇOIS. 2012. MENTAL FILES. Oxford: Oxford University Press.

RECANATI, FRANÇOIS. 2016. MENTAL FILES IN FLUX. Oxford: Oxford University Press.

RUSSELL, BERTRAND. 1911. *Knowledge by Acquaintance and Knowledge by Description.* PROCEEDINGS OF THE ARISTOTELIAN SOCIETY 11: 108–128.

SCANLON, T.M. 1998. WHAT WE OWE TO EACH OTHER. Cambridge MA: Harvard University Press.

SCHROETER, LAURA. 2007. *The Illusion of Transparency.* AUSTRALASIAN JOURNAL OF PHILOSOPHY 85: 597–618.

SCHROETER, LAURA. 2008. *Why Be an Anti-Individualist?* PHILOSOPHY AND PHENOMENOLOGICAL RESEARCH 77: 105–141.

SCHROETER, LAURA. 2012. *Bootstrapping Our Way to Samesaying.* SYNTHÈSE 189: 177–197.

SCHROETER, LAURA. 2013. *Are Concepts Creatures of Darkness? Critical Discussion of Robert Stalnaker's* Our Knowledge of the Internal World. ANALYTIC PHILOSOPHY 54: 277–292.

SCHROETER, LAURA & FRANÇOIS SCHROETER. 2009. *A Third Way in Metaethics.* NOÛS 43: 1–30.

SCHROETER, LAURA & FRANÇOIS SCHROETER. 2014. *Normative Concepts: A Connectedness Model.* Philosophers' IMPRINT 14: 1–26.

SCHROETER, LAURA & FRANÇOIS SCHROETER. 2015. *Rationalizing Self-Interpretation.* In PALGRAVE HANDBOOK OF PHILOSOPHICAL METHODS, ed. CHRISTOPHER DALY. Basingstoke: Palgrave Macmillan.

SCHROETER, LAURA & FRANÇOIS SCHROETER. 2016. *Semantic Deference vs. Semantic Coordination.* AMERICAN PHILOSPHICAL QUARTERLY 53: 193–210.

SIMESTER, A.P. 2000. *Can Negligence Be Culpable?* In OXFORD ESSAYS IN JURISPRUDENCE, 4th series, ed. JEREMY HORDER. Oxford: Oxford University Press.

SIMESTER, A.P. 2013. *A Disintegrated Theory of Culpability.* In THE SANCTITY OF LIFE AND THE CRIMIAL LAW: THE LEGACY OF GLANVILLE WILLIAMS, ed. DENNIS J. BAKER & JEREMY HORDER, 178–203. Cambridge: Cambridge University Press.

SINGER, PETER. 1972. *Famine, Affluence, and Morality.* PHILOSOPHY & PUBLIC AFFAIRS 1: 229–243.

STAVROPOULOS, NICOS. 1996. OBJECTIVITY IN LAW. Oxford: Oxford University Press.

STEVENS, ROBERT. 2007. TORTS AND RIGHTS. Oxford: Oxford University Press.

STRAWSON, P.F. 1974. SUBJECT AND PREDICATE IN LOGIC AND GRAMMAR. London: Methuen.

TOH, KEVIN. 2005. *Hart's Expressivism and His Benthamite Project.* LEGAL THEORY 11: 75–123.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063

Toh, Kevin. 2011. *Legal Judgments as Plural Acceptances of Norms.* In Oxford Studies in Philosophy of Law 1, ed. Leslie Green & Brian Leiter, 103–137. Oxford: Oxford University Press.

Toh, Kevin. 2019. *Legal Philosophy à la carte.* In Dimensions of Normativity: New Essays on Metaethics and Jurisprudence, ed. David Plunkett, Scott Shapiro & Kevin Toh, 221–247. Oxford: Oxford University Press.

Williamson, Timothy. 2006. *Conceptual Truth.* Aristotelian Society supp. 80: 1–41.

Zack, Naomi. 2002. Philosophy of Science and Race. New York: Routledge.

Downloaded from https://www.cambridge.org/core. University of Melbourne Library, on 04 Sep 2020 at 01:53:58, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S1352325220000063