**Concepts as shared regulative ideals**

Laura Schroeter and François Schroeter

What is it to share the same concept? The question is an important one since sharing the same concept explains our ability to non-accidentally coordinate on the same topic over time and between individuals. Moreover, concept identity grounds key logical relations among thought contents such as samesaying, contradiction, validity, and entailment. Finally, an account of concept identity is crucial to explaining and justifying epistemic efforts to better understand the precise contents of our thoughts. In general, concepts provide a stable framework that allow us to keep track of specific topics in thought and talk, despite open-ended variability in individuals’ understanding of the topic in question. The key question, then, is what psychological and social facts could play these roles? Elsewhere, we have argued for a specific relational model of concept identity, the *connectedness* account (e.g. Schroeter 2012, Schroeter and Schroeter 2014, 2015). Our aim in this chapter is to further explain the motivations behind our account, to address worries about transitivity and vagueness, and to contrast our approach with closely related accounts of concept identity developed by François Recanati and Simon Prosser.

What’s distinctive of our approach is that we seek to vindicate the first-person epistemic perspective of concept users. Concepts, on our account, play a crucial normative role in setting *regulative ideals* for the representational practices in which individual subjects participate. This focus on the normative role of concepts – as opposed to a purely causal explanatory role – motivates our approach to concept identity and our toleration of vagueness and borderline cases.

**1. Keeping track of a topic**

We take concepts to be *basic ways of keeping track of a topic in thought and talk*. Concepts, on this approach, reflect the ways that we amass and store information about a single topic over time, so that the accumulated information is ready to be redeployed in recognition, induction, and inference. And at the interpersonal level, shared concepts reflect our ways of coordinating on the same topic in language.

Our aim in this chapter is to articulate an account of concept identity that reflects the first-person perspective of concept users. Many theorists focus on the predictive and explanatory role of concepts in characterizing how a person’s reasoning unfolds. But this is not the perspective that thinkers adopt when deploying and fine-tuning their own representational states. From the first-person perspective, there is an important distinction between our actual understanding of a topic and the real nature of that topic: we could be ignorant or mistaken about the precise content of our thoughts. And similarly, there is an important distinction between our actual understanding of the logical/conceptual relations among our thoughts and their real logical relations: the logical structure of our thoughts is not always transparent to us.

To get an intuitive feel for the role of concepts in thought, consider a mundane example. Suppose you see a photo of Emmanuel Macron in his office, and this sets off a train of thought:

(M) *Macron is the youngest French head of state since Napoleon. I wonder if Macron’s shorter than Napoleon. Look, Macron has a copy of Le Rouge et Le Noir on his desk – something odd there. He clearly thinks he’s a modern-day Julien Sorel. Well, he did marry his high school drama coach. I hope his story has a happier ending.*

This example highlights the cognitive phenomenology characteristic of concepts. Throughout this train of thought, your thoughts are immediately presented to your conscious attention as obviouslypertaining to the same topic: *Macron*. When you’re entertaining these thoughts, moreover, the question of sameness of topic seems closed: it seems rationally incontrovertible that *Macron is Macron*. Indeed, it seems like there is no evidence that could be relevant to justifying or reinforcing this apparently trivial identity claim: sameness of topic seems epistemically basic. We call this implicit cognitive phenomenology linking your thoughts *the appearance of de jure sameness* of topic (ADJS). Thoughts connected via ADJS contrast with cases where sameness of topic seems open to rational doubt: e.g. you might recall a childhood acquaintance, *Emmanuel*, who you strongly suspect is Emmanuel Macron. Even if you become convinced that young Emmanuel is in fact Emmanuel Macron, the identity will not strike you as rationally incontrovertible and epistemically basic the way the thoughts in (M) do.

We take the epistemic features we’ve identified as ADJS to constitute the *epistemic signature* *of concept identity* (Schroeter 2012). When you deploy the same concept in thought, sameness of topic seems:

1. Obvious: the elements of thought are automatically treated as pertaining to the same topic – no reasoning is required.
2. Rationally incontrovertible: from the thinker’s current perspective, sameness of topic appears rationally guaranteed: it is impossible for the thinker to imagine circumstances where a≠a.
3. Epistemically basic: there is no way to reinforce the justification for a=a: any evidence cited will appear less certain than the identity claim itself.

In effect, then, the question of sameness seems closed while you’re actually engaged in a train of thought. However, we want to stress that this cognitive phenomenology is a mere appearance, which disappears if you adopt a meta-level perspective on your own thoughts. If you ask whether *this* thought was about the same person as *that* thought, it no longer seems obvious, incontrovertible, and epistemically basic that your thoughts are about the same thing. This fact, together with epistemic basicness, strongly suggests that ADJS is generated by subpersonal cognitive mechanisms for binding and storing information. And we think the metaphor of a *mental file folder* is an apt way of characterizing the structure of the relevant subpersonal binding mechanisms involved in generating these person-level appearances. We’ll return to these suggestions about cognition below.

Concepts mark *formal logical relations* among thoughts. They determine whether token representational states – token elements of a particular thought – belong to the same semantic type. Thus, the epistemic signature of ADJS constitutes an individual’s implicit awareness of these formal relations. An individual’s grasp of this formal structure can be contrasted with their *substantive understanding* of the topic picked out by their thoughts. Your substantive understanding of a topic like *Macron*, for instance, includes your implicit recognitional, inferential, emotional or motivational dispositions, as well as a suite of consciously accessible, conceptually articulated beliefs, desires, fears, and intentions. Your substantive understanding associated with your concept is constantly changing, as you update your attitudes or modify your implicit dispositions through experience and reasoning. However, the conceptual structure of your standing attitudes seems to remain fixed throughout its existence.

Our focus in this chapter will be on *strict concept identity*. Strict concepts are representational state types that respect the following two constraints:

1. *Cognitive Difference Principle* (CDP): If two elements of thought are not linked by ADJS, then they deploy different concepts.
2. *Semantic Difference Principle* (SDP): If two elements of thought differ in their reference, then they do not express the same strict concept.

The first constraint is determined by the distinctive epistemic signature of concept identity, ADJS. With a nod to Frege, we call this the Cognitive Difference Principle since it captures the rational relations that Frege used to individuate distinct senses. We take CDP to be a general constraint on concept identity. The second constraint, the Semantic Difference Principle, is distinctive of the subclass of *strict* concepts. Singular concepts are paradigms of strict concepts: whenever your concept MACRON is deployed in a thought, the concept picks out the very same thing, the individual human being Macron, as the reference. So if an element of thought refers to something besides Macron, it must express a distinct strict concept. We take strict concepts to be ways of keeping track of a particular object, kind, or property, which helps to fix the truth conditions of thoughts in which the concept occurs. Other paradigm strict concepts may be certain natural kind concepts (e.g. GOLD), normative kinds (e.g. MORALLY RIGHT), or mathematical terms (e.g. THREE).[[1]](#footnote-1)

Strict concepts are philosophically important because they ground logical relations, such as samesaying, contradiction, entailment, and validity.[[2]](#footnote-2) These logical relations require that the content remains stable between different deployments of the concept. And they also require that this sameness of content should be subjectively obvious to the thinker in the ways that are distinctive of ADJS. If we want to explain what makes it the case that an individual’s thoughts stand in strict logical relations, then, we need a theoretical account of strict concept identity which explains how concepts are related to that individual’s cognitive states. In particular, this account should explain the following two features:

* Reference (semantic content): how does the reference of a thought relate to the individual’s implicit understanding of its reference (i.e. application and inferential dispositions governing that token representational state)?
* Logical relations: how do real logical relations relate to the thinker’s implicit understanding of those relations (via ADJS)?

In our view, the subject’s current understanding of the precise reference of their thoughts and of the logical relations they stand in are both reliable guides to the facts. But they are fallible.

Before we turn to theoretical accounts of strict concepts, we would like to highlight three further constraints on concept identity.

1. Stability (of concepts): concepts normally remain stable *over time* and *between interlocutors*.
2. Flexibility (of understanding): concepts can remain stable through open-ended variation in substantive understanding of their contents.
3. Justifiability (of content): conceptual competence puts one in a position to get closer to the truth about the precise contents of one’s own thoughts, given relevant information about one’s circumstances and ideal reflection.

The Stability constraint is a natural consequence of the idea that concepts are ways of keeping track of a topic in thought and talk. If we are to accumulate and store bodies of information *as* about the same topic, then concepts must remain stable over time and between individuals. The process of accumulating information will naturally lead to variation in individuals’ substantive understanding of the topic picked out. The Flexibility constraint acknowledges the range of variation in substantive understanding compatible with apparent *de jure* co-reference over time and between individuals. A sceptic can reject almost any specific assumption about Macron – say, that he’s president of France or that he’s a human being – without undermining the appearance of direct logical coordination with the rest of us. Finally, the Justifiability constraint ties the correct semantic interpretation of an individual’s words and thoughts to their own reflective perspective. Intuitively, a semantic assignment that flouts an individual’s own ideal standards for epistemic self-correction would be uncharitable. However, Justifiability merely requires that the individual’s epistemic standards be reliable – not that they be perfectly accurate. Since we are trying to characterize individuals’ first-person perspective on their own concepts, we take Justifiability to be a plausible constraint on semantic interpretation.

Together, these three constraints reflect the role of concepts in grounding rational inquiry and debate about the nature of a familiar topic. Concepts should remain stable through rational inquiry about the nature of the topic, they should allow for open-ended variation in our substantive commitments about the topic, and they should put us in a position to come to a reliable verdict about the nature of that topic through inquiry. These constraints (iii-v) thus reflect the general epistemic role of concepts, while the first two constraints (i-ii) reflect the role that strict concepts play in grounding strict logical relations.[[3]](#footnote-3)

**2. Representational traditions**

Let’s now consider how these constraints can be met by a theoretical account of *strict concept identity*. There are two broad approaches to explaining what makes it the case that two token elements of thought express the same concept. The traditional Fregean approach focuses on the patterns of substantive understanding associated with token elements of thought, whereas the more recent relational approach focuses on subpersonal cognitive binding mechanisms that generate a distinctive person-level cognitive phenomenology.[[4]](#footnote-4)

1. Matching model: Two token elements of thought express the same concept iff each token is associated with a matching pattern of substantive understanding of the content (e.g. specific application dispositions, inferential dispositions, mental models, etc.), and this match:
2. explains ADJS between the tokens, and
3. guarantees sameness of content (in matching contexts).
4. Relational model: Subpersonal cognitive relations bind token elements of thought into distinct semantic units (e.g. mental files, representational traditions), and these binding relations:
5. explain ADJS between the tokens, and
6. define the default units for semantic interpretation (which explain sameness of content in normal circumstances).

The two approaches offer contrasting explanations of CDP and SDP. On the matching model, the basic explanatory units are token elements of thought, each of which has its own proprietary pattern of understanding and its independently determined content. On the relational model, in contrast, the basic explanatory units are networks of tokens, demarcated by subpersonal binding relations. Both models can explain why ADJS is a necessary condition for sharing a concept, and how sameness of semantic content depends on purely internal cognitive states. On both accounts, however, strict concept identity depends in part on normal external circumstances to secure sameness of content.

We have argued elsewhere against the matching approach to concept identity: in particular, we claim that it fails to meet the Flexibility constraint and that it posits implausible psychological mechanisms for keeping track of a topic in thought.[[5]](#footnote-5) However, we won’t rehearse those arguments here. Instead, we will sketch the version of the relational model we favour, with a special emphasis on how the account reflects individuals’ first-person epistemic perspective on their own thoughts.

On our account, the basic unit for semantic interpretation is a *representational tradition* (Schroeter and Schroeter 2014). Phenomenologically, a representational tradition is demarcated by synchronic, diachronic and interpersonal chains of ADJS relations. But this person-level cognitive phenomenology, we suggest, is generated by subpersonal cognitive binding mechanisms linking token representational states. At a time, an individual’s thoughts are organised by a mental filing system:

* + *Mental files at a time*: An individual’s standing attitudes and cognitive dispositions are bound by subpersonal cognitive mechanisms, mental files, in such a way as to generate ADJS relations among the bound elements.

For instance, consider the reasoning episode (M): the elements of occurrent thoughts that you’d express with ‘Macron’ are all linked via ADJS relations. On our account, these appearances are explained by a stable subpersonal filing system, which binds not just the thoughts that occur in (M) but a bundle of accumulated attitudes and dispositions that you are disposed to treat as pertaining to the same topic – *Macron*. Over time, the standing states bound by your ‘Macron’ file will gradually change, as you learn more about the topic and revise or forget some of your earlier views. But at each stage, a mental file binds memories and standing attitudes inherited from prior stages. The overlapping chains of standing attitudes bound by a file over time demarcate a continuous *representational tradition*.

* + *Diachronic representational traditions*: The standing attitudes in a file are linked via subpersonal cognitive mechanisms that generate ADJS relations to past thought episodes, either directly (via standing attitudes) or indirectly (via continuous chains of standing attitudes).

From the first-person perspective, the activation of a standing memory generates ADJS to the past state from which it was derived. Diachronic representational traditions are demarcated by the underlying cognitive mechanisms that generate these appearances.

There is a similar cognitive binding relation linking your attitudes with those of others in your community via words in a public language. In normal circumstances, your automatic linguistic parsing mechanisms will directly link an interlocutor’s use of ‘Macron’ to your own mental file, in such a way as to generate ADJS. As long as you remain focused on the object-level contents conveyed, it will seem obvious, rationally incontrovertible, and epistemically basic that there’s one topic in question, *Macron*. (But of course, that appearance will disappear as soon as you ascend to the meta-level perspective – just as it does when we reflect on our own thoughts.) This interpersonal ADJS relation can be used to define shared representational traditions:

* + *Shared representational traditions*: An individual’s mental files are linked via subpersonal linguistic parsing mechanisms that generate ADJS to others’ mental files, either directly (via direct communication) or indirectly (via continuous chains of direct communication).

From the first-person perspective, using the same words in conversation seems to guarantee that we’re coordinating on the same topic. So shared representational traditions can be traced from our direct interlocutors outwards into the community, as well as backwards into their past.

Since our account seeks to capture the first-person perspective on concepts, we want to stress that ordinary thinkers are implicitly aware of their own representational traditions. You are normally aware that you have been keeping track of a topic like *Macron* over time, gradually accumulating new information about it. And you are alive to the possibility that your current stock of assumptions may be partly ignorant or mistaken about the topic picked out. This implicit meta-level awareness informs your epistemic efforts to stay on topic, correct incoherencies in your understanding over time, and seek to fill gaps in your current understanding. You are also aware that you are coordinating with others in accumulating information about the topic, as you accept testimony and engage in joint inquiry and debate. And if you come to a surprising conclusion about the topic, e.g. that Macron doesn’t really exist, or that he’s really an American stooge, you don’t just take this conclusion to show that your own current MACRON attitudes are ill-founded: you’ll take your conclusion to apply to your past MACRON thoughts and those of your linguistic community. You’re committed to the idea that we’ve all been mistaken all along about Macron. That commitment reflects your implicit meta-cognitive awareness of our shared representational tradition.

**3. Interpretation and disambiguation**  
The second aspect of an account of strict concept identity is a theory of semantic interpretation. On our account, representational traditions form the *default units* for interpretation. That is, we start from the common sense presumption that we have all been thinking and talking about the same topic all along when we use expressions like ‘Macron’, and we look for a semantic interpretation that can vindicate that presumption. Of course, there is likely to be variation in opinions about the precise nature of the reference (or other content) picked out – and each of us is epistemically fallible. The key question, then, is: how do we discover the truth about the content of a shared representational tradition, given this disagreement? And can we assign a univocal semantic content to the tradition as a whole? According to SDP, sameness of strict concept entails sameness of semantic content.

Our account of the semantic interpretation of a representational tradition appeals to norms of *rationalizing self-interpretation* (Schroeter and Schroeter 2014, 2015). This self-interpretive method involves two distinct aspects:

* + 1. *Representational interests*: identify the stable representational interests that have sustained and justified the shared representational tradition.
    2. *Best satisfier*: identify the semantic contents (if any) that best meet the interests at stake in the representational tradition.

To get a sense of how these steps work, consider the case of ‘free will’. We have a shared representational tradition of distinguishing between actions that are freely chosen as opposed to performed under duress, and we hold each other accountable for acts done of one’s own free will. But what exactly does it take for an action to count as an instance of ‘free will’? The first step on our account is to identify the *point* of categorizing actions as ‘free’. Some theorists hold that the dominant interest at stake hinges on our interest in *causal explanation*: free actions are those whose causal source is the agent herself as opposed to forces outside the agent. Others argue that our dominant interest in classifying actions as free is in *assigning praise and blame*: agents are held responsible for their free actions. These alternatives are partly normative judgments about which aspects of our categorizing practices have been most important to justifying our shared representational tradition. And the correct answer will depend in part on empirical facts – on historical and social facts about our classificatory practices and on objective facts about the actions we classify. The second step of rationalizing interpretation assigns a reference (or other semantic content) that best meets these interests, given the empirical facts. For instance, if our dominant interest is in assigning praise and blame, then an interpreter will need to decide which empirical facts about an action make it accountable to praise or blame. Perhaps free actions, on this interpretation, might require some reflective capacities on the part of the agent, and a lack of certain psychological pressures.

We hope the example makes it clear that rationalizing self-interpretation of shared representational traditions allows for open-ended inquiry and debate about both steps of interpretation. Our account outlines the underlying structure of these debates, but does not seek to provide an algorithm for resolving them. We believe the structure we’ve outlined – identifying the representational interests at stake in a shared representational practice and identifying a content that best satisfies those interests – will strike philosophers as a plausible description of our own methods for defending philosophical analyses of familiar terms. Our claim is that this method is an *idealization of our* *best epistemic methods* for getting closer to the truth about what exactly we’ve all been talking about all along (Schroeter and Schroeter 2015).

In contrast to the traditional matching approach, a relational account of concepts makes sharing a concept easy – no matching patterns of substantive understanding are required. Instead, *we bootstrap our way to de jure semantic coordination* (Schroeter 2012, Schroeter and Schroeter 2015):

Bootstrapping: rationalizing self-interpretation seeks to vindicate the appearance of *de jure* sameness (ADJS) in a tradition by identifying a univocal interpretation for the shared tradition.

We start the process of semantic interpretation with the default presumption that the appearance of *de jure* sameness is likely to be veridical, and our methods for self-interpretation seek to vindicate that presumption. The aim of interpretation is to find a semantic interpretation that is mutually justifiable to the participants in the tradition – not just to an individual thinker. On this approach, *de jure* semantic coordination does not require a precise match in individuals’ reference-fixing criteria prior to inquiry. Instead, semantic coordination is explained by the fact that participants in a shared tradition look for mutually justifiable interpretations of their words. This self-interpretive disposition helps to bridge the gap between variations in individuals’ initial substantive understanding of a word: the fact that we’re *trying* to coordinate on a common interpretation helps make it the case that we *do* coordinate.

However, we need one further condition in our account in order to secure a univocal reference for strict concepts. The problem is that our representational traditions can involve *undetected semantic drift* over time and across a linguistic community. Consider Gareth Evans’ ‘Madagascar’ case: the name was originally used for a region of the African mainland, the story goes, but Marco Polo mistakenly took it to refer to the large island off the coast and this latter understanding has now come to be dominant (Evans 1973). There was no break in the representational tradition: there is a continuous chain of ADJS relations linking our current use of the name back to the original users in Africa. But intuitively the reference of the representational tradition with ‘Madagascar’ has changed: it originally referred unambiguously to the mainland, and now refers unambiguously to the island. This sort of semantic drift is ubiquitous in natural language.

A further interpretative step is required to handle such cases. The first point to note is that that there are two incompatible but well-supported referential interpretations of the tradition. Second, these two interpretations are well-suited to the interests of differently situated individuals within that tradition: earlier thinkers were collecting information about the mainland, whereas later speakers were oriented towards the island. We suggest that in this sort of case, rationalizing self-interpretation should seek to *partition* the representational tradition in such a way that participants’ commitment to keeping track of a single location comes out as *true* *within their partition*. Thus, in the case of strict concepts there is a third stage in rationalizing self-interpretation:

1. *Disambiguation*: when there are incompatible but equally good interpretations of a representational tradition, rationalizing self-interpretation seeks to partition the tradition so that a single semantic content can be assigned to each partition.

This final stage of interpretation allows us to vindicate the claim that reference has shifted over time. And it captures our reflective epistemic commitments about which beliefs should be accepted as true, and which should be rejected as false.[[6]](#footnote-6)

With this account of representational traditions and rationalizing self-interpretation in hand, we can now characterize strict concept identity in terms of *ideally disambiguated representational traditions*. To a first approximation, we can individuate disambiguated traditions as follows:

Strict concept identity: two token elements of thought instantiate the same strict concept iff:

(*Connectedness*) they are connected to each other via a continuous representational tradition (constituted by ADJS relations);

(*Congruence*) the understanding and historical context associated with the tokens does not diverge so radically as to undermine univocal semantic interpretation; and

(*Univocal Reference*) the representational interests at stake in the tradition require rationalizing interpretation to assign a univocal referential content.

The Connectedness constraint grounds concept identity in causal-historical ADJS relations between token elements thought, while the Congruence constraint requires that the linked tokens be apt for a univocal semantic interpretation if they are to count as instantiating the same concept. If not, the tradition will be subject to disambiguation. The last constraint, Univocal Reference, is distinctive of strict concepts: it requires that all tokens of the same strict concept co-refer. This rules out semantic interpretations of a shared tradition that assign context-sensitive referential contents.[[7]](#footnote-7)

In Evans’ ‘Madagascar’ case, for instance, there is a continuous representational tradition of ADJS relations connecting your contemporary token use of that name with uses of a similar-sounding name by 12th century African speakers (possibly ‘Mogadishu’). So the early and late uses satisfy the Connectedness constraint on concept identity. However, these token uses do not satisfy the Congruence constraint. The patterns of understanding and use associated with early and late phases of the representational tradition are so radically divergent that they cannot be assigned the same reference.[[8]](#footnote-8) Instead, rationalizing interpretation will seek to partition the representational tradition in such a way that uses in the early phase of the tradition univocally refer to Mogadishu and late uses univocally refer to Madagascar. So the two token uses in question do not express the same concept. The token uses that fall within one of these partitions, however, will satisfy the constraints on strict concept identity: connectedness, congruence, and univocal reference.

This account meets our logical and epistemic desiderata on a theory of strict concept identity. First, it secures *logical relations* among token deployments of a concept, by satisfying both CDP and SDP. CDP is satisfied since token elements of thought that are *not* connected via ADJS express distinct concepts.[[9]](#footnote-9) And SDP is satisfied since a difference in reference (semantic content) entails a difference in strict concept. The account is also well suited to filling the role of concepts in grounding rational inquiry and debate. Sharing a strict concept on this account is easy: all that’s required is to join a shared representational tradition with univocal representational purport and that one’s understanding not diverge too radically from that of other participants. But there is no specific pattern of substantive understanding required for conceptual competence. Thus Stability and Flexibility are easily met by disambiguated traditions. And Justifiability is also neatly explained, since semantic contents are assigned on the basis of an idealization of the very self-interpretative methods that rational epistemic agents hold themselves accountable to. In sum, our relational model can explain strict logical relations among thoughts without constraining the scope of rational inquiry and debate.

**4. Concepts as regulative ideals**

In the previous section, we suggested concepts should be individuated in terms of disambiguated representational traditions. But what *kind* *of thing* are concepts so conceived? If you are used to thinking of concepts as mental representations that play a particular type of causal role in explaining individuals’ online reasoning, then our account of concept identity may strike you as a category mistake. How can an ideally disambiguated representational tradition be part of a *causal explanation of an individual’s actual reasoning process* in coming to accept a particular attitude?

Our response is that we are focusing on the role of concepts from the first-person perspective of concept users *as epistemic agents*. When we use a concept – or reflect on our own conceptual practices – we are not primarily focused on causal explanation of our own reasoning. Rather, we take our concepts to fix the *truth conditions* of our thoughts and their *logical relations* at a time, over time, and between individuals. Moreover, we take our current understanding to be fallible: each of us could be ignorant or mistaken about the reference or the logical properties of our own words and thoughts. From our first-person perspective, concepts fix epistemic ideals for how we *should* understand token elements of thought. Concepts are *shared regulative ideals* governing our representational states.

Let’s unpack that claim. When we say that strict concepts are *ideals*, we mean that the semantic contents and *de jure* sameness relations of token elements of thought are determined by *idealization* of our epistemic practices. On our account, the correct semantic interpretation is determined by an idealization of the self-interpretive process among participants in a representational tradition. It is *not* determined by an individual’s actual understanding (either implicit cognitive dispositions or explicit reflective verdicts), since by the individual’s own lights this could be mistaken. This fallibility holds not just for our understanding of the referential content of a thought, but also for its logical relations to other thoughts. The appearance of *de jure* sameness is not a failsafe guide to the reality of *de jure* sameness, any more than our current categorizing dispositions are a failsafe guide to how an object, kind, or property is really individuated. Just as a correct semantic assignment depends in part on ideal, empirically informed rationalizing interpretation, a correct partition of a representational tradition depends on ideal, empirically informed disambiguation. It is this idealized disambiguation that determines logical relations like samesaying, entailment, and contradiction.

Next, when we say these semantic and logical ideals marked by strict concepts are *regulative*, we mean that they provide *normative standards of correctness* for an individual subject’s practices of epistemic reflection on their attitudes. First, in fixing reference, strict concepts determine precisely what we are thinking about when we use terms like ‘Madagascar’ or ‘free will’. They thereby determine the correctness conditions – objective truth-conditions – of the thoughts in which those concepts figure. These semantic facts, in turn, determine when it’s semantically correct to accept, reject, or refine our mental states. Second, in marking disambiguated *de jure* sameness relations, strict concepts set standards of correctness for the logical coherence of our beliefs: they help determine which *sets of beliefs* (or intentions) can coherently be accepted together. If our representational tradition with ‘Madagascar’ has conflated two distinct places, for instance, we will need to correct our beliefs about the logical relations between past and present claims. Thus, strict concepts play a regulative role by setting standards of correctness that determine the acceptability of both (i) *specific beliefs* (and dispositions)as true characterizations of the topic and (ii) *sets of attitudes* as logically consistent perspectives on that topic. They set objective normative standards for how rational epistemic agents *should* regulate their own representational traditions.

Our approach makes concepts *normative kinds*: concepts set normative standards of correctness for the acceptance of attitudes and cognitive dispositions. As we noted earlier, this approach contrasts with the dominant approach to concept individuation, which treats concepts as causal-explanatory kinds. On a causal-explanatory approach, the primary theoretical role of concepts is to mark causal-explanatory patterns that are useful in prediction and explanation of individuals’ online reasoning. An account of concept identity that reflects how we actually tend to reason will be ill-suited to capturing our reflective epistemic commitments about how we ought to reason. It is this normative perspective, we believe, that fits most comfortably with our first-person perspective on our own representational traditions.

Although we take the dominant theoretical role of strict concepts to be as regulative ideals, we want to emphasize that these ideals reflect causal patterns within shared representational practices. Strict concept identity, after all, is grounded in a complex set of empirical facts. Representational traditions are demarcated by a variety of subpersonal binding mechanisms that link token representational states at a time (via mental files), over time (via stable attitudes within files), and between individuals (via automatic linguistic parsing). These representational traditions serve as conduits for sharing information at a time, over time and between individuals. So traditions will tend to generate broadly similar patterns of associated understanding – especially when participants in the traditions are gathering information from the same environment. This similarity in understanding, moreover, is regulated through social practices and coordinated epistemic methods of holding each other accountable to a common standard of correctness. So while strict concepts are individuated with an eye to their normative role, they nonetheless mark broad empirical patterns that are important from the point of view cooperative epistemic agents.

In the next two sections of this chapter, we’d like to clarify our approach to strict concepts by addressing two important objections. First, our account fails to vindicate individuals’ transparent epistemic access to the sameness or difference of their own concepts. Second, our account faces challenges in vindicating the transitivity of strict concept identity. We’ll take these concerns in turn.

**5. Transparency**

Many philosophers believe that the transparency of concept identity is a strict constraint on an adequate account of concepts. The idea is that sameness or difference in the concepts deployed in thought will be *subjectively obvious* to the thinker – at least when those thoughts occur together in the same episode of reasoning.

For instance, Michael Dummett writes:

It is an undeniable feature of the notion of meaning–obscure as that notion is–that meaning is *transparent* in the sense that, if someone attaches a meaning to each of two words, he must know whether these meanings are the same. (Dummett 1978)

Dummett is expressing the widespread view that it is epistemically obvious to a thinker whether their words express the same meaning (concept). More specifically, we suggest, the intuition behind Dummett’s claim is that the epistemic signature of concept identity, ADJS, is an *infallible* indicator of sameness concept within an individual’s own thoughts at a given time:

Synchronic Transparency: Two co-conscious elements of thought deploy the same concept iff they are linked by ADJS.

This is not to say that thinkers cannot make a mistake about concept identity *in their online reasoning* – but any such mistake will involve a breakdown in minimal rational functioning. As Tyler Burge puts it, “brute empirical mistakes” about the sameness or difference of the concepts one employed within a single episode of reasoning are impossible: errors about redeploying the same concept must be explained by a *rational* *failure* of carelessness, malfunction, or irrationality.[[10]](#footnote-10)

Paul Boghossian sums up the motivations for this view nicely:

We…ascribe thoughts to a person…for two related purposes; on the one hand, to enable assessments of his rationality and, on the other, to explain his behavior. As these matters are currently conceived, a thought must be epistemically transparent if it is to play these roles. Without transparency, our conceptions of rationality and rational explanation yield absurd results. (Boghossian 1994)[[11]](#footnote-11)

The idea, then, is that concepts mark *standards of minimal rationality*. To reason in ways that flout the appearance of obvious and incontrovertible sameness (ADJS) would constitute a breakdown in minimal rational functioning in the thinker. Moreover, since a breakdown of minimal rationality is extremely rare, we can use the ADJS relations demarcated by an individual’s mental files at a time to causally *predict and explain* a thinker’s online reasoning.

To illustrate the commitments of our account of concepts with respect to transparency, let’s consider a slightly modified version of a slow-switching case sketched by Boghossian. Consider Peter, who has been unwittingly shifted from Earth to Twin Earth many years ago. In his youth on Earth, Peter once glimpsed Pavarotti floating on the water of Lake Taupo – an exciting encounter that became a treasured memory. Subsequently, he was surreptitiously transported to Twin Earth and continued his life there, never suspecting his switch. After many years of interacting with XYZ and following the news about opera stars, Peter eventually acquires the ability to form beliefs about XYZ and Twin-Pavarotti. One day Peter sees Twin Pavarotti floating on the waters of his hotel pool. This leads him to recall his earlier encounter and draw a conclusion about the coincidence. Peter might express this line of reasoning as follows:

NP (*Naive Pavarotti reasoning*)

1. Perceptual belief: ‘Pavarotti is floating on water now (in the pool).’
2. Standing memory: ‘Pavarotti floated on the water then (at Lake Taupo).’
3. Inferential belief: ‘So Pavarotti is doing the same thing now as then.’

The reasoning seems perfectly valid from Peter’s point of view. But on our account of concepts as disambiguated representational traditions, Peter’s reasoning will have the following conceptual form:

DINP: *Disambiguating interpretation of NP*

1. Perceptual Belief: PAVAROTTI2 FLOATS ON WATER2 AT T2.
2. Standing memory: PAVAROTTI1 FLOATS ON WATER1 AT T1.
3. Inferential belief: [???] DOES THE SAME THING AT T1 AND T2.

Peter has one continuous representational tradition associated with the term ‘Pavarotti’, but because of his switch there are two equally plausible referential candidates for this tradition: the earthly man, Pavarotti1, and his twin-earthly counterpart, Pavarotti2. Just as with the Madagascar case, ideal disambiguation will seek to partition this tradition into distinct sub-traditions, in such a way that Peter’s standing memory is grouped with other token thoughts as referring to *Pavarotti1* and the perceptual belief is grouped with token thoughts as referring to *Pavarotti2*. (*Mutatis mutandis* for ‘water’.) The inferential belief (III), we suggest, will have indeterminate content. On this interpretation, Peter’s reasoning is based on conceptual equivocation: it trades on the misleading appearance of *de jure* sameness due to the fact that distinct strict concepts are included in a single representational tradition. Moreover, there is no way for Peter to detect this mistake without learning about his switch. This is precisely the sort of brute empirical error that Synchronic Transparency rules out.

Other anti-individualists have sought to avoid that result (Burge 1998, Recanati 2016, Prosser 2019). These theorists propose strategies for vindicating a version of the transparency thesis by reinterpreting the semantic contents of Peter’s attitudes. The most common anti-individualist approach is to hold that Peter’s mental files gradually shift their reference over time, so that Peter’s synchronic deployments of a file within an occurrent episode of reasoning must be assigned the same interpretation. On this account, the concepts involved in Peter’s standing attitudes gradually shift over time, due to the drift in the semantic contents of his mental files: either the reference becomes *indeterminate* over time (Recanati 2016:131), or it may *shift* to twin-earthly things (Prosser 2019: 476). On Prosser’s view, Peter’s original belief PAVAROTTI1 FLOATS ON WATER1 AT T1 is transformed over time into a belief that PAVAROTTI2 FLOATS ON WATER2 AT T1. On this account, Peter’s reasoning in NP is valid, since it deploys the same (Twin-Earthly) concepts throughout. But NP is unsound, since the standing attitude (ii) is false. In contrast, Recanati takes all the deployments of ‘Pavarotti’ and ‘water’ in NP to have indeterminate content. On this account, NP is valid but suffers from reference failure. On either account, the appearance of *de jure* sameness linking an individual’s thoughts within any single episode of reasoning is guaranteed to be veridical. So Synchronic Transparency is satisfied.[[12]](#footnote-12)

Saving Synchronic Transparency, however, comes at a cost: Peter has lost direct epistemic access to his earlier thoughts, since his standing memories now employ different concepts with different contents. This means that Peter’s standing attitudes do not preserve their logical form over time. Insisting on Peter’s transparent access to logical relations *at a time* forces one to construe Peter as rationally malfunctioning *over time* (through no fault of his own). More generally, Recanati and Prosser restrict strict concept identity to synchronic intrapersonal thoughts – or to thoughts that occur close enough together in time as to preclude referential drift.[[13]](#footnote-13) On these accounts, then, there are no logical relations connecting attitudes over any significant stretch of time within an individual or over social distance within a community.

In contrast, our account privileges diachronic over synchronic rationality. We take this approach to be a natural consequence of our starting position of privileging an epistemic agent’s perspective on their own thoughts over the causal-explanatory perspective of an external observer. Let’s unpack these ideas.

As epistemic agents, we have an interest in accumulating and storing information derived from the recurring objects, kinds, and properties in our environment, which can then be used in induction and reasoning. This is what our mental filing system allows us to do: it provides the cognitive infrastructure for binding and storing evolving bodies of attitudes and dispositions *as* pertaining to a single topic. Up until his switch, Peter’s ‘Pavarotti’ file binds a body of beliefs and cognitive dispositions derived from the singer Pavarotti1 which reliably track the facts about him. So Peter’s mental filing system at that time is operating in roughly the way it should: gradually accumulating standing attitudes and cognitive dispositions and storing them *as* pertaining to a stable topic, Pavarotti1.

The first point to emphasize is that this filing practice is an essentially *diachronic* process: the goal is to accurately track and accumulate information about a feature of the world to use in later induction and reasoning. Moreover, an attitude’s maintaining the same semantic content over time is central to the well-functioning of a filing system. This stability allows the standing attitudes to preserve their warrant and truthover time.

To illustrate these points, consider Peter’s standing attitudes before the switch. Peter’s standing belief (ii) was originally formed when he glimpsed Pavarotti1 at Lake Taupo. We can assume that Peter was a responsible epistemic agent, and his belief was both true and warranted at the time he formed it. Subsequently, his standing belief was periodically reactivated, for instance when reminiscing about the famous singer and daydreaming about revisiting New Zealand in hopes of another chance encounter. In these circumstances, it’s uncontroversial that (ii) maintains the same content over time up until his switch. And assuming that Peter has not acquired new evidence that threatens to defeat or undermine his original perceptual justification, his warrant for the belief is also preserved.

Let’s now consider what happens after the switch. According to defenders of Synchronic Transparency, the content of Peter’s standing belief (ii) changes. After the switch, on these accounts, Peter’s standing belief about his encounter with Pavarotti1 undergoes gradual referential drift as he interacts with his new environment. After a suitable period, (ii) acquires *indeterminate* *semantic content* (as Recanati suggests):

(II') Standing memory: PAVAROTTI1-cum-PAVAROTTI2 FLOATED ON WATER1-cum-WATER2 AT T1.

And perhaps after a long period on Twin Earth (ii) will eventually acquire determinate twin-earthly contents (as Prosser suggests):

(II'') Standing memory: PAVAROTTI2 FLOATED ON WATER2 AT T1.

On either interpretation, the shift in content means that Peter’s standing belief (ii) is no longer true. Moreover, the shift in the concepts involved in that persisting attitude entail that Peter’s diachronic reasoning, from (II) to (II') or (II"), is no longer valid, and that his original epistemic warrant for his persisting standing belief no longer holds. On interpretation (II"), his past perceptual encounter with Pavarotti1 and his efforts to confirm what happened at that time *no longer provide any justification* for his standing belief about Pavarotti2. Thus, the standing belief is cut off from its original evidential base. And on interpretation (II'), his epistemic warrant is in an even worse state, since there is simply no empirical evidence that could support a belief about the indeterminate object, *Pavarotti1-cum-Pavarotti2*. Without content preservation warrant is not preserved. So Peter’s carefully collected and confirmed standing belief (ii) is now epistemically worthless.

In addition, if these interpretive verdicts are taken as a regulative ideal for Peter’s epistemic self-governance, then Peter should now reject his standing belief (ii). But following this advice would significantly impoverish Peter’s accumulated store of worldly information. According to interpretation (II'), Peter should reject *every* attitude and disposition bound by his ‘Pavarotti’ file, since they have all become referentially confused and thus unwarranted. And according to interpretation (II''), Peter should reject all the assumptions *acquired before his switch*, since those standing commitments are about Pavarotti2, with whom Peter was not acquainted at the time. So if Peter uses interpretations (II') or (II'') as a guide for regulating his standing attitudes, large swathes of attitudes and dispositions which were painstakingly accumulated over time must be rejected as unwarranted. Peter’s ‘Pavarotti’ file will either be emptied of information, or significantly impoverished.

Our approach to concept identity avoids these consequences. A disambiguating interpretation allows Peter’s standing attitudes dating from Earth to retain their original semantic contents over time. So they remain true and retain their original epistemic warrant. Construed as a regulative ideal, our account’s advice for Peter is *not* that he should abandon large swathes of his ‘Pavarotti’ attitudes, but rather that he should disambiguate his filing system, so as to segregate his Earthly from his Twin Earthly beliefs. Disambiguating self-interpretation reconstrues the logical form of the naive train of reasoning as DINP. By interpreting (ii) as (II), our account vindicates Peter’s *diachronic rationality*: Peter has a standing belief with a stable semantic content and stable sources of warrant over time. Correcting his reasoning in NP does not require Peter to reject his warranted belief (II), but to change his disposition to draw the conclusion (III). More specifically, Peter should reject the ADJS linking his PAVAROTTI1 and PAVAROTTI2 beliefs as a *false appearance*, whose default warrant is undermined by his new empirical evidence about his switch. A more thorough disambiguation of his ‘Pavarotti’ attitudes will partition his thoughts into new files that reflect the real co-reference relations among his thoughts (according to norms for charitable self-interpretation). Our account thus provides advice that both preserves warranted standing attitudes and prevents equivocal reasoning in the future.

One might worry, however, that (II) construes Peter as *synchronically* *irrational* when he trades on the presumed identity of Earthly and Twin Earthly things in NP. Our interpretation DINP construes Peter as making a *basic logical mistake*, by conflating two distinct concepts. And one might object that this interpretation is intuitively implausible. Peter’s mistake is not due to a failure in *minimal rational coherence*: his mistake is not due to craziness or crossed wires. Rather, his mistake is due to an empirical error: he fails to detect his switch to a new environment. So Peter seems *rationally blameless* in drawing the inferences that he does prior to learning about his switch. Yet according to DINP, Peter fails to recognise that he’s using distinct concepts – so he’s failing minimal standards of logical coherence.

In response, we’d like to make two points. First, we agree with proponents of Synchronic Transparency that Peter is rationally blameless. Peter conforms to the standards of *minimal procedural rationality* required for normal cognitive functioning when he relies on ADJS as a guide to the real *de jure* sameness relations. So Peter is minimally rational in virtue of trading on ADJS. Second, our interpretation of the conceptual structure of Peter’s thoughts, DINP, does *not* entail that Peter is violating these norms of minimal procedural rationality – that he’s cognitively malfunctioning, careless, or crazy. Concepts, on our account, are not individuated so as to mark minimal procedural rationality. Instead, they mark the regulative ideals that rational epistemic agents should hold themselves accountable to over time. In Peter’s case, we need to choose between an account of concept individuation that reflects the fact that, (a) as a synchronic agent, Peter is minimally procedurally rational when he draws the inference NP, or (b) as a diachronic agent, Peter is epistemically warranted in maintaining his standing attitudes over time. Our account favours (b) when the two principles conflict, since we believe this better reflects the regulative ideals to which epistemic agents should hold themselves accountable. So on our account, it’s possible to conflate distinct concepts (and thus make a logical error of equivocation in reasoning) due to brute empirical ignorance. This is because the role of concepts, on our account, is to mark the ways our actual filing systems shouldbe refined to reflect the warrant structure of our beliefs, not the way our filing systems are currently organised.

Our position, moreover, has the significant advantage of vindicating the idea that our thoughts can stand in direct logical relations over time and between individuals. Normally, Peter’s thoughts employ the very same strict concepts over time and Peter shares those concepts with his linguistic community. A relational account that respects Synchronic Transparency must abandon these claims to Stability – one of our initial desiderata on an account of concept individuation (in §1). The stability of concepts is central to our perspective as rational epistemic agents on our own thought contents. We start with the presumption that all of us have been thinking and talking about the same things all along when we use terms like ‘Pavarotti’ or ‘water’. And this presumption of *de jure* sameness of content leads us to treat others’ understanding and use of these terms as relevant to our own inquiry and debate about the precise nature of those things. In short, concept stability is deeply engrained in our epistemic norms for rational inquiry and debate. Shared concepts allow us to put our heads together and identify mutually relevant features of our environment as the reference of our words and thoughts.

In closing, we’d like to stress that our argument in this section is *not* that our interpretation DINP better fits with common sense intuitions about the switching case than the alternatives proposed by Recanati, Prosser, and Burge. We think common sense intuitions are malleable and they can shift depending on our theoretical interests. Instead, our claim is that our account of disambiguating interpretation is superior for a specific *theoretical purpose*: it is better suited to vindicating the way rational epistemic agents regulate their own representational states. Our account sets an appropriate *regulative ideal* for concept users themselves. It is consistent with our position that other ways of individuating representational state types might be superior for other theoretical purposes. So we can afford to be ecumenical about the different ways of individuating fine-grained representational state types that might deserve the name ‘concept’. Perhaps our shared representational tradition of talking about ‘concepts’ itself stands in need of disambiguation. For the purposes of causal explanation of an individual’s online reasoning, it may be useful to appeal to an account that identifies strict concepts with temporal stages of individuals’ mental file, as Recanati and Prosser suggest. These explanatory accounts of strict concept identity are in a sense complementary to ours, in that they posit the same underlying cognitive structures, but focus on different elements of that structure as deserving the name ‘concept’. File-stages capture the *causal mechanisms* that ground an individual’s actual reasoning dispositions at a time: an individual’s dispositions to trade on identity (via ADJS) and their current understanding of the content. In contrast, our account is built to capture the truth-conditions and logical relations that rational epistemic agents *should* hold themselves accountable to: shared conceptual traditions (individuated in part by ADJS) and rationalizing interpretation of those shared traditions. Ideally disambiguated traditions, we suggest, capture the *normative standards* that are relevant for epistemic self-governance, rather than actual patterns of understanding. Insofar as you’re interested in concepts as setting epistemic standards of correctness for belief and reasoning, our account is to be preferred.

1. **Transitivity and semantic drift**

Let’s turn to another important worry generated by cases of referential drift. Strict concept identity is supposed to be an *equivalence relation*: reflexive, symmetric, and transitive. Our strategy for securing an equivalence relation hinges on our metasemantic theory: disambiguating interpretation seeks to partition representational traditions into discrete equivalence classes, in which the token elements of thought are not just connected by ADJS but can also be assigned the same semantic content (according to norms of rationalizing self-interpretation). However, inevitably there will be some tokens within a tradition that cannot be assigned a determinate univocal reference through rationalizing disambiguation. But if interpretations generated by disambiguation include some indeterminate semantic assignments, this will yield partitions with vague borders. This sort of vagueness, however, undermines the claim that strict concept identity is a genuine equivalence relation, since the relation will not be transitive.

The worry can be sharpened by using referential drift to construct a sorites paradox for our account of strict concept identity. Normally, the patterns of understanding and history of use associated with a new token within an individual’s representational tradition will differ in very minor respects from that of immediately prior tokens. In cases like Peter’s, there is simply the gradual accretion of a history of interacting with and (mis)classifying a new object (Twin Pavarotti) as ‘Pavarotti’ within his ongoing representational tradition. From one moment to the next, such incremental historical changes do not seem to warrant any shift in the interpretation of Peter’s new perceptual beliefs. Designating one time as the instant where the reference of Peter’s new ‘Pavarotti’ thoughts shifts from Pavarotti to Twin Pavarotti would be hopelessly *ad hoc* and unjustifiable. However, after ten years, it’s uncontroversial that a semantic change has occurred, and thus that Peter’s will unwittingly be deploying a new concept Pavarotti2 in his interactions with Twin Pavarotti.

Peter’s case thus fits the sorites schema:

Pα1

∀t(Pαt→Pαt+1) (Tolerance)

–––––––––

∀t(Pαt)

An element of thought α1 in Peter’s ‘Pavarotti’ tradition at time 1 has the property P of picking out Pavarotti as the reference. According to the principle of Tolerance, the incremental difference between Peter’s state at any two adjacent times, t and t+1, does not justify the assignment of a different reference. It follows, then, that a change of reference is impossible: at all times, Peter’s ‘Pavarotti’ thoughts in the tradition have the same reference (and express the same concept). So according to the common sense interpretive principle of Tolerance, referential drift is impossible. And yet, as Peter’s case illustrates, common sense interpretation is nonetheless committed to the possibility of referential drift over a sufficiently long time. Our task, then, is to explain how conceiving of strict concept identity as an equivalence relation can be reconciled with the phenomenon of vague borderline cases.

Our answer hinges on the role of concepts as *regulative ideals*. Our account of concept identity is defined in terms of an idealization of our epistemic practices: ideal rationalizing and disambiguating interpretation of a representational tradition seeks to demarcate strict concept identity as an equivalence relation that respects both CDP and SDP. However, this ideal is not fully realizable, given the fine-grained variability within our representational traditions and the vagueness inherent in interpreting the different interests at stake in our representational traditions. Sometimes, these idealizations will fail to give clear advice about the reference and logical relations that determine which attitudes (or sets of attitudes) are correct (and consistent). These cases fall into the vague penumbra of strict concept identity. However, our suggestion is that this vagueness at the borders poses no threat to the reality of strict concepts as empirical representational state types. Nor does it undercut the usefulness of the idealizations involved in treating concept identity as demarcating strict equivalence classes.

Consider the role of idealization in science. As Angela Potochnik notes, scientific idealizations typically involve *false assumptions* about empirical systems:

Idealizations are assumptions made without regard for whether they are true and often with full knowledge they are false. (Potochnik 2020)

Let’s look at the example of ideal gas laws. The idealization presupposed by these laws is the transparently false assumption that gas molecules are non-interacting point particles. This transparent falsehood is epistemically valuable, Potochnik argues, because it allows us to abstract away from competing causal aspects within a physical system, such as the different sizes of specific gas molecules and different gravitational forces exerted on molecules in the system. Because of this idealization, ideal gas laws have a limited range of application (they fail to hold at very high or low temperatures), and the laws admit of exceptions even within their core range of application. However, Potochnik argues that the false idealization is nonetheless epistemically valuable in its own right. By neglecting orthogonal causal forces, we are better able to recognize and understand *real causal patterns* that hold across a broad range of conditions. These causal patterns are real aspects of the empirical world that allow us to causally manipulate, predict, and explain changes to those systems. Moreover, the idealization makes these patterns *cognitively accessible* to us in ways that would be impossible if we were to try to understand the full complexity of forces involved in particular systems – or if we were to simply rely on more accurate mathematical models.[[14]](#footnote-14)

One might worry that false idealizations have the potential to mislead us into making false predictions about particular systems. But there are ways of strategically managing that risk. Mark Wilson has long argued that scientists have a wealth of *ad hoc* strategies for avoiding scientific errors suggested by the misleading logical properties of their theoretical models (Wilson 2017, 2006, 1994). Often we learn by experience to avoid drawing inferences suggested by our ideal models, implicitly coming to recognize boundary conditions for their application. This move might be an initial stage in the process of developing a more comprehensive theory of the causal patterns involved. But in many cases, there simply is no way of fitting all the practical know-how of applying the model into a manageable theoretical description: our best way of understanding complex systems is to hone our pragmatic strategies for managing our application of relatively simple idealized models (Wilson 2017).

We suggest that this approach to idealizations in the philosophy of science provides an apt model for the idealizations involved in concept identity. Just as the idealization of gas molecules as point particles *simplifies* and *misrepresents* the complex gaseous systems they model by ignoring competing forces and boundary conditions, the idealization of concepts as sharply delimited state types *simplifies* and *misrepresents* the complexity of the psychological systems they model by ignoring vagueness. In both cases, the idealization abstracts away from certain empirical details of actual physical systems that obscure important overall patterns.

Let’s unpack this analogy. Idealizations are not perfectly accurate representations of the empirical phenomena they depict: they abstract away from certain empirical details in order to clearly represent recurring patterns. So which empirical details are glossed over? And why is this idealization empirically and normatively useful?

Our account of strict concept identity abstracts from the vagueness inherent in demarcating representational state types within actual psychological systems. The Univocality constraint requires that all tokens of the same strict concept must have precisely the same semantic content. Thus, disambiguating interpretation seeks to partition representational traditions into discrete equivalence classes in such a way that all members of the resulting partitions can be rationally interpreted as having precisely the same content. Thus, according to the model, all tokens within a tradition stand in determinate relations of (veridical) *de jure* sameness to a specific subset of thoughts within that tradition. The model thus eliminates any vagueness in reference or concept identity. However, this idealization runs into problems when we consider the complex empirical facts that support semantic interpretation. The patterns of understanding and history of use within a tradition come in indefinitely many variations and gradations over time and between individuals. Similarly, the representational interests at stake in a tradition will vary smoothly over time and across different social contexts. And the number of the ADJS connections linking an individual’s files to a shared representational tradition will also vary. This complexity and smooth variation in the empirical facts that ground rationalizing self-interpretation will inevitably lead to vague penumbral cases when seeking to demarcate precise boundaries for disambiguated traditions. The intermediate stages of inadvertent referential drift are a case in point. In Peter’s case, for instance, it’s intuitively clear that his initial ‘Pavarotti’ thoughts pick out Pavarotti, and that eventually at least some of his ‘Pavarotti’ thoughts pick out Twin-Pavarotti. But between these two extremes, there will be a period where the competing interpretations are *on a par*: the empirical facts about Peter’s representational tradition during this time do not determinately mandate one interpretation over the other. In such cases, it is impossible to draw a perfectly sharp empirical cut-off point for strict concept identity in a non-ad hoc way. The idealized model thus *misrepresents* the vague empirical boundary conditions for strict concept identity.[[15]](#footnote-15)

So if interpreted literally, the model is false. Nonetheless, we maintain that this idealized model of strict concept identity captures empirical and normative regularities, which are important from the point of view of concept users.

First, like the ideal gas laws, the idealization of strict concept identity highlights *real empirical patterns* that are explanatorily important. According to the idealization, disambiguating interpretation partitions our actual representational traditions in such a way as to respect the ideal epistemic commitments of participants in those traditions. So strict concepts mark idealized partitions that our thoughts *should* respect given ideal rationality and full empirical information. As we’ve stressed, this idealized model is inaccurate as a characterization of our *actual* mental filing practices in a number of important ways. For instance, we sometimes systematically conflate distinct individuals, kinds, or properties within a single file or tradition, and our representational traditions are subject to gradual referential drift due to our partial ignorance about the topic picked out. Nonetheless, our actual representational traditions tend to approximate the ideal of univocal representational traditions, since we are normally pretty good at keeping track of a topic via our representational traditions. Conflation, drift, and borderline cases of *de jure* sameness are the exception rather than the rule. As a consequence, the idealization of perfectly sharp univocal concepts tracks significant *empirical regularities* in our actual representational practices.

Moreover, the model tracks *causal mechanisms* that are relevant to explaining successes and to detecting errors in the ways we implicitly keep track of a topic over time and between individuals. Our account of strict concept identity is based on subjectively salient empirical facts about our representational practices, which figure in our epistemic practices of self-governance. These include:

1. *The cognitive phenomenology* of ADJS, linking individual thoughts at a time, over time, and between individuals through subpersonal binding mechanisms,
2. *Shared representational traditions* defined by these ADJS relations, linking variable patterns of substantive understanding and use over time and between individuals,
3. *Epistemic dispositions* to police and correct our representational practices through object-level inquiry and debate, and
4. *Reflective* *interpretive* *dispositions* to correct our representational practices through meta-level self-interpretation and disambiguation.

The idealized model of disambiguated representational traditions highlights the importance of these mechanisms in regulating our representational practices, and abstracts away from the messy fine-grained empirical details involved in borderline cases. The idealization of sharp conceptual boundaries is a simple and explanatorily powerful model of real causal mechanisms. And it provides a cognitively efficient way to keep track of these normatively relevant causal patterns.

The second aspect of why the idealization of perfectly sharp concept identity is valuable is the *practical role* it plays in guiding our reflective epistemic practices. Conceiving of disambiguated traditions as equivalence classes with perfectly precise borders is a useful regulative ideal for managing our representational practices. Presupposing sharp conceptual boundaries, with no vague penumbras, encourages us to see ourselves as participating in shared representational practices that preserve strict logical relations of *de jure* sameness of topic. This model vindicates our naïve epistemic practices of collecting information *as* about a single stable topic over time despite variations in understanding, sharing this collected information through testimony, and engaging in joint inquiry and debate about the nature of that topic. The model also encourages us to *look for* a single determinate content for the tradition as a whole when we engage in object-level inquiry or meta-level self-interpretation.

And when disambiguating confused representational traditions, the model encourages us to find partitions that minimize semantic indeterminacy. Thus, the idealized model promotes epistemic practices that scaffold, regulate, and reinforce increased determinacy within our empirical representational traditions. It is a simple and powerful way of understanding our own naïve conceptual practices, which fosters epistemically rational methods for correcting and refining those practices.

In sum, we have argued that the idealized model of strict concepts as lacking vague borderline cases is false, but that the false idealization it is both epistemically and practically valuable. The model captures important empirical and normative patterns that would be obscured by a more accurate model.

In general, causal-historically individuated kinds generate borderline cases of kind identity. Just as there are borderline cases of species identity when a biological lineage transitions from one stable genotype to another, there are borderline cases of concept identity when a representational tradition drifts from one stable semantic content to another. Nonetheless, transitivity of (kind) identity holds for these kinds, provided we avoid including cases that fall within the transitional periods. So one way to avoid drawing false conclusions from the idealization of positing an equivalence relation is to simply avoid drawing transitivity inferences that extend across these transitional phases in a causal-historical tradition.

Of course, we can’t completely avoid characterizing an individual’s beliefs *during* the transitional phase when a representational tradition is undergoing semantic drift. In Peter’s case, for instance, there will be a transitional period where the justification for interpreting Peter’s newly formed beliefs as referring to the Earthly Pavarotti will be on a par with the justification for interpreting it as about Twin Pavarotti. During this transitional period, there is no way to justify a unique semantic content for Peter’s newly formed beliefs – and thus no way to assign them to a single disambiguated tradition. For these beliefs, there are no determinate facts about semantic content and strict content identity.

Nonetheless, we still need to *talk* about such cases in everyday interactions. So how are we to characterize the contents of these thoughts and their logical relations? We suggest that interpreters treat transitional cases in an *ad hoc* and context-sensitive way, depending on the interpreter’s own explanatory interests. For instance, we may be interested in highlighting the continuity of Peter’s borderline thoughts with his later Twin earthly beliefs, or the apparent logical validity of his inferences, or the epistemic warrant or correctness of forming new beliefs in this representational tradition. Within a conversational context, we can find ways of highlighting these different aspects of Peter’s transitional beliefs using our own disambiguated traditions. But as long as we are sensitive to the fact that these thoughts are vague borderline cases, we can avoid drawing incorrect inferences based on concept identity. Fortunately, we are implicitly sensitive to situations like Peter’s where the semantic content of a representational tradition is in the process of transitioning from one content to another. In such cases, we simply avoid relying on our idealized model of sharp conceptual boundaries.

In sum, risks of taking the idealization of strict concept identity too literally in reasoning can be managed through canny context-dependent epistemic strategies. This allows us to reap the epistemic and normative benefits of the idealization without the costs: the model highlights the empirical regularities and normative standards that are important from the perspective of thinkers themselves in managing their representational practices.

**7. Conclusion**

Throughout this chapter, our approach to concepts has been guided by the first-person perspective of concept users on their own representational states. From this engaged perspective, concepts are ways of keeping track of a topic that mark logical relations among token thoughts at a time, over time, and between individuals. This perspective led us to construe concepts as *regulative ideals* for rational self-governance, grounded in rationalizing interpretation of our shared representational traditions. This perspective also led us to the conclusion that concepts should be individuated in a way that privileges diachronic rationality over synchronic rationality. From the first-person perspective, semantic drift can lead us to conflate distinct concepts in our reasoning: the ultimate truth conditions and logical relations that we *should* hold our thoughts accountable to may depend on empirical facts about which we are ignorant or mistaken. Finally, we suggested that conceiving of concepts as regulative ideals provides a new way of resolving the tension between strict concept identity (construed as an equivalence relation) and the intransitivity and vagueness generated by semantic drift. If concepts are idealizations, we can be relaxed about vague borderline cases in ways that would be inappropriate if we were to treat concepts as precise representations of empirical reality. Nonetheless, like scientific idealizations, the idealizations involved in strict concept identity capture real and important features of our representational practices.

Finally, in an ecumenical spirit, we want to reiterate that our account is just one way of individuating fine-grained representational state types that might deserve the name ‘concept’. In particular, we conceded that other, causal-explanatory interests may lead one to look for accounts of concept identity that more closely track individual’s actual cognitive organization. In particular, file-stage theories like those of Burge, Recanati, and Prosser share much of the cognitive infrastructure posited by our own account of concept identity. These approaches may be preferable for the purposes of causal explanation of an individual’s reasoning at a time. But we believe such accounts lead to a distortion of the normative standards of correctness governing both the *truth conditions* of token thoughts and the *logical relations* among them. Insofar as philosophers want an account of concepts that sets appropriate normative standards for belief, we suggest that the connectedness account is best suited to vindicate the verdicts generated by our best empirically informed epistemic practices.[[16]](#footnote-16)

**References**

Boghossian, P. (1994). "The Transparency of Mental Content." *Philosophical Perspectives* 8: 73-122.

Burge, T. (1982). Other Bodies. *Thought and Object*. A. Woodfield. Oxford, Oxford University Press**:** 97–120.

--- (1988). "Individualism and Self-Knowledge." *Journal of Philosophy* 85: 649–663.

--- (1998). Memory and Self-Knowledge. *Externalism and Self-Knowledge*. P. Ludlow and N. Martin. Palo Alto, CA, CSLI Publications**:** 351-371.

Cumming, S. (2013a). "Creatures of Darkness." *Analytic Philosophy* 54: 379–400.

--- (2013b). "From Coordination to Content." *Philosophers' Imprint* 13: 1–17.

Dummett, M. (1978). *Truth and Other Enigmas*. London, Duckworth.

Evans, G. (1973). "The Causal Theory of Names." *Proceedings of the Aristotelian Society* Supp. 47: 187–208.

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

Onofri, A. (2017). "The Publicity of Thought." *Philosophical Quarterly* 68: 521–541.

Potochnik, A. (2020). "Idealization and Many Aims." *Philosophy of Science* 87: 933–943.

Prosser, S. (2019). "Shared Modes of Presentation." *Mind and Language* 34: 465–482.

Recanati, F. (2016). *Mental Files in Flux*. Oxford, Oxford University Press.

--- (forthcoming). Sameness of Mode of Presentation. *Sharing Thoughts*. M. Valente and V. Verdejo, Oxford University Press.

Schroeter, L. (2007). "The Illusion of Transparency." *Australasian Journal of Philosophy* 85: 597-618.

--- (2008). "Why Be an Anti-Individualist?" *Philosophy and Phenomenological Research* 77: 105-141.

--- (2012). "Bootstrapping our Way to Samesaying." *Synthese* 189: 177–197.

Schroeter, L. and F. Schroeter (2014). "Normative Concepts: A Connectedness Model." *Philosophers' Imprint* 14: 1-26.

--- (2015). Rationalizing Self-Interpretation. *Palgrave Handbook of Philosophical Methods*. C. Daly. Basingstoke, Palgrave Macmillan.

--- (2018). "Keeping Track of What's Right." *Canadian Journal of Philosophy* 48: 489-509.

Valente, M. and A. Onofri (2022). "A Puzzle about Communication." *Review of Philosophy and Psychology* Online prepublication: <https://doi.org/10.1007/s13164-021-00606-w>.

Wilson, M. (1994). "Can We Trust Logical Form?" *Journal of Philosophy* 91: 519–544.

--- (2006). *Wandering Significance*. Oxford, Oxford University Press.

--- (2017). Pragmatics’ Place at the Table: Lessons from Multiscalar Science. *Physics Avoidance*. Oxford, Oxford University Press**:** 1–50.

1. We have formulated strict SDP in terms of *reference*, given that our main focus in this chapter will be on concepts that pick out determinate referential contents. However, the notion of a strict concept can be extended to include other types of contents, such as truth-functions or purely expressive contents for normative terms. To accommodate these additional types of strict concepts, SDP would need to be reformulated in terms of *semantic content* rather than *referential content*. The key point is that strict concepts contribute to the correctness conditions of the thoughts in which they figure. Thus the meanings of context-sensitive expressions like ‘here’ or ‘tall’ and polysemous expressions like ‘book’, whose content shifts depending on the context of use, do *not* express the same strict concept. These words may, however, express representational state types that meet the other desiderata on concept identity: CDP and (iii)-(v) below. We call such representational states *relaxed concepts.*  [↑](#footnote-ref-1)
2. We use the term ‘samesaying’ to mark the opposite of a direct logical contradiction – it involves two thoughts with the same logical form, or two elements of thought of the same semantic type. [↑](#footnote-ref-2)
3. It’s worth stressing that our motivation for positing shared concept is *not* that they are required in order to explain interpersonal communication. As Matheus Valente and Andrea Onofri (2022) forcefully argue, interpersonal communication and understanding can be achieved without positing precisely matching semantic contents. They rightly note that proponents of shared contents between individuals must provide an alternative motivation for interpersonally shared concepts. Our response to this challenge is that shared concepts are required to vindicate logical relations among thoughts within an individual over time and between individuals. Shared concepts are also needed to vindicate coordinated rational inquiry and debate about precise semantic contents of our thoughts. [↑](#footnote-ref-3)
4. Here we part company with theorists (Fiengo and May 2006, Onofri 2017) who seek to account for semantic coordination via *beliefs* about co-referring thoughts. Our account is instead based on the cognitive phenomenologyof ADJS,which we take to be irreducible to belief. [↑](#footnote-ref-4)
5. See, for instance, (Schroeter 2012, Schroeter and Schroeter 2014). [↑](#footnote-ref-5)
6. Disambiguation can also impose partitions across a linguistic community *at a time* to account for divergent local dialects; and it can impose partitions *within an individual idiolect at a time* to disambiguate systematic conflations. For more on disambiguation, see (Schroeter 2008, 2007, Schroeter and Schroeter 2014). Disambiguation will not fully eliminate indeterminacy in concept identity conditions: there can attitudes that cannot be assigned exclusively to one partition. This is an issue we will return to below. [↑](#footnote-ref-6)
7. For a discussion of conceptual practices that constitute univocal representational purport, see (Schroeter and Schroeter 2018); for the congruence constraint, see (Schroeter and Schroeter 2014). The formulation in the text is only a first approximation because it fails to address a further complication for relational accounts: fusion or mistaken fission within a representational tradition. Consider Kripke’s Paderewski case: there is a univocal public tradition of using ‘Paderewski’ for a single man (a pianist turned politician), but Peter mistakenly takes there to be two individuals named ‘Paderewski’ (a pianist and a politician) whom he keeps track of via distinct mental files. Each of Peter’s ‘Paderewski’ files satisfy (i)-(iii). So on our current formulation, both would count as instantiating the same strict concept. But according to CDP, Peter’s two files instantiate distinct concepts. This would lead to a failure of transitivity, so that strict concept identity would not be an equivalence relation. We take this to be a compelling objection to the sufficiency of (i)-(iii). What’s missing is a further constraint barring fusion and illicit fission in univocal representational traditions. In our view, Peter is not fully competent with the shared public ‘Paderewski’ concept precisely because he fails to grasp the *logical relations* among thought contents that shared concepts are supposed to mark. He fails to understand, for instance, that what he learns about Paderewski’s pianistic career is logically related to what he learns about Paderewski’s political career. Formulating this further constraint on strict concept identity is a delicate matter, which we won’t take up here. But see (Cumming 2013a, 2013b) for a detailed account of a semantic coordination constraint for the Paderewski case, which is used to develop a coordination-based account of fine-grained contents. [↑](#footnote-ref-7)
8. It’s worth emphasizing that both the medieval and contemporary tokens in the tradition are *understood* as having a univocal reference. That is, the thinker’s substantive understanding of the token expression is governed by epistemic dispositions for collecting and correcting information *about a specific location*. Given these epistemic dispositions, rationalizing interpretation should seek to assign a univocal reference. In particular, it’s not plausible to treat medieval and contemporary users of token names as sharing a stable, but context-variable meaning. [↑](#footnote-ref-8)
9. It’s worth stressing, however, that the ADJS relations may be indirect. So in *diachronic* or *interpersonal* cases, it may not be subjectively obvious that we are deploying the same concept. Matching models of concept identity also have this consequence. [↑](#footnote-ref-9)
10. (Burge 1988). For a survey of different approaches to transparency, and a contrast of our approach with Burge’s, see (Schroeter 2007). [↑](#footnote-ref-10)
11. Boghossian notes that his transparency constraint is in the spirit of Frege and Russell, who both took epistemic access to the contents of one’s words to be central. Boghossian’s formulation of the motivation for transparency echoes that of Tyler Burge, whose anti-individualist views he is criticizing in the paper. According to Burge, concepts and finely individuated thought contents are “the stuff of which explanations of [an individual’s] actions and assessments of his rationality are made” (Burge 1982). [↑](#footnote-ref-11)
12. Tyler Burge’s account is slightly different in structure. According to Burge, standing states will normally preserve their contents when reactivated at later times. But Burge holds that this default must be overridden to preserve Synchronic Transparency: the ADJS links to prior thoughts within a specific train of reasoning, he says, “will often take precedence” in determining the content of an activation of a standing attitude (1998: 360). So on Burge’s account ADJS links within conscious episodes of reasoning are guaranteed to be veridical, but standing attitudes don’t have determinate semantic contents or determinate logical relations independently of the context of use. See (Schroeter 2007) for a critical discussion of Burge’s view. [↑](#footnote-ref-12)
13. Recanati (forthcoming) suggests that Transparency is ‘fundamentally synchronic’, since inference always takes place at a particular time and within a particular thinker (Ms:15-16). Prosser (2019: 476) defends Transparency over short intervals of time: ‘Over small enough intervals, there is no determinate shift in reference between tokens, given that the intervals can be arbitrarily small.’ [↑](#footnote-ref-13)
14. “[I]n my view, idealization can play an epistemically crucial role. Idealized representation enables epistemic agents to grasp a causal pattern that a more accurate representation of the phenomenon would obliterate. […] [B]ecause of the cognitive value of simple patterns, scientists regularly choose to sacrifice some accuracy of those phenomena to the end of grasping the patterns the phenomena embody.” (Potochnik 2020: 942) [↑](#footnote-ref-14)
15. The problem here is not confined to relational accounts of strict concept identity – it also arises for matching models. On a broadly Fregean approach, for instance, sharing the same concept is a matter of token elements of thought being associated with *matching patterns of understanding*, such as matching criteria for identifying the reference or matching motivational roles. These conditions can be captured by an abstract definition, which may seem to demarcate perfectly sharp boundaries for sharing a concept. But this appearance quickly evaporates when we ask precisely which cognitive states suffice for a thinkers’ implicit grasp of that definition. An individual’s understanding of a word like ‘gold’ is governed by complex layers of dispositions and cognitive structures – including dispositions to recognise instances on the basis of different sorts of evidence or testimony, inferential dispositions that reflect idiosyncratic theoretical beliefs, and epistemic dispositions to correct these first-order dispositions on the basis of specific methodological commitments. All of these dispositions come in varying strengths, and their precise structure depends on the specific range of mental files and perceptual capacities an individual thinker possesses. Given this complexity, how exactly can we set a principled cut-off point for sharing precisely the same psychological pattern of understanding? Any sharp psychological cut-off point, we submit, will be arbitrary. [↑](#footnote-ref-15)
16. For helpful feedback, we’d like to thank the editors, anonymous reviewers, and participants in the 2021 Sharing Thoughts Workshop and at the Dianoia Institute’s seminar series at the Australian Catholic University. [↑](#footnote-ref-16)