THE EMERGENCE OF MENTAL CONTENT: AN ESSAY IN THE METAPHYSICS

OF MIND

Philip Woodward

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements for the degree

Doctor of Philosophy

in the Department of Philosophy,

Indiana University

August 2015

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the

requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____

Timothy O'Connor, Ph.D. (Chair)

_____

Kirk Ludwig, Ph.D.

_____

Frederick Schmitt, Ph.D.

_____

Colin Allen, Ph.D.

July 29, 2015

For my parents

Mary Christina Lamkin Woodward

and

Dr. William Henry Woodward, Jr.

to whom I owe more than I can say

All of us are aware that our perceptual, imaginative and cognitive states can be *directed at* or *about* things. We can have perceptual experiences as of the clouds; we can undergo imaginative episodes as of riding a roller coaster; we can entertain thoughts about where to dedicate our financial resources. The phenomenon of mental directedness or aboutness which these examples illustrate is familiar and ubiquitous, yet at the same time it has a number of mysterious features. For one thing, if "aboutness" is a relation, one of its relata can be non-existent: I can think about mythical beasts and historical figures as readily as I can think about the person sitting across from me. For another, it is not clear how to locate the "aboutness" relation within the structure of the natural world. A mental state of mine can be about my mother. What would it be for a state of the weather (a downpour, say) to be about my mother? It is hard to know what that could mean. And if we can make little sense of a meteorological state's being about my mother, can we understand how a *neurological state* could be about my mother?

In the endeavor to explore these puzzles and others like them, philosophers have introduced a few technical terms. The phenomenon of directedness or aboutness they have dubbed "intentionality". (The term descends from a Latin word that means *to aim at*.) That toward which a mental state is directed is its "intentional content." (The technical phrase 'intentional content' expresses a mixed metaphor—an *aiming*-metaphor mixed with a *containment*-metaphor—thus capturing another puzzling feature of the phenomenon. Content is in some sense internal to a mental state—it is that which a mental state contains—while at the same time being in some sense external to a mental

state—it is that toward which a mental state points.) In the following pages I assume that the reader is familiar with these terms and with some of the historical discussion around them.

I frame my discussion as an inquiry into *intentional properties*. Intentional properties are those features of mental states in virtue of which such states have the intentional content they have. My dissertation is about the metaphysics of intentional properties: what mental states have to be, what their nature must include, in order for such states to have intentional contents. In particular, my dissertation explores a consciousness-based metaphysics of intentional properties.

Central to the dissertation is the "Phenomenal Grounding Thesis", or PGT: *all instances of intentional properties are partly grounded in instances of phenomenal properties* (where phenomenal properties are the felt qualities of conscious experience). The dissertation divides roughly into two parts. In the first part (chapters 1 and 2 and half of chapter 3), I provide arguments in support of PGT. In the second part (the other half of chapter 3, and then chapters 4-6), I attempt to explain how phenomenal properties could ground intentional properties.

I don't argue directly for PGT. But I try to motivate it by arguing against leading rival theories of intentional properties in **chapter 1** ("The Acquaintance Argument for Intrinsic Intentionality") and then by arguing for a weaker, modal claim, in **chapter 2** (**"**Motivations for the Phenomenal Grounding Thesis"). The rival theories in question are so-called "naturalistic" theories. Rather than grounding intentionality in consciousness, these theories seek to ground intentionality in causation, broadly construed (including not only token causal relations but causal dispositions and nomic relations). My objection to

these theories is rooted in an observation about the epistemic role that conscious

intentional states play. A subject's semantic self-knowledge—her knowledge of the

content of her mental states—explains her ability to recognize when two mental states

have content that is the same or different. This wouldn't be possible unless intentional

properties were available as the objects of introspective acquaintance; but because we can

only be introspectively acquainted with the intrinsic features of our mental states, and

because all naturalistic theories treat intentional properties as non-intrinsic, naturalistic

theories are not able to account for the sort of semantic self-knowledge that explains our

introspective discriminatory capacities—or so I argue. Now, it is open to the naturalizer

to deny that our discriminatory capacities are explained by semantic self-knowledge (but

to hold instead that those capacities are brute); so saying is not only phenomenologically

implausible but also explanatorily inelegant—it treats as an explanans a set of phenomena

that ought to be treated as the explanandum.

Such is my argument against rival theories. In **chapter 2** I continue to try to

motivate PGT by reviewing and building upon a number of arguments in the literature

that purport to reveal a close modal connection between intentionality and consciousness:

first, that some phenomenal states are metaphysically sufficient for the obtaining of

intentional states (a claim I call 'Phenomenal Sufficiency'); and second, that some

phenomenal states are metaphysically necessary for the obtaining of intentional states (a

claim I call 'Phenomenal Necessity'). While not all of the arguments I canvass are

equally persuasive, the cumulative case for both claims is strong.

To argue for PGT, motivating it isn't enough; I also need to defend it, because

both common sense and cognitive science point to *non-conscious* intentional states, and

it's natural to think that non-conscious intentional states pose a direct challenge to PGT. I argue in **chapter 3** ("A Theory of Phenomenal Grounding") that this is not so. Purported cases of non-conscious content can be accommodated within my framework, if we treat some of them as fictitious but useful (a natural way to go with some of the posits of cognitive science), if we treat some of them as dispositional (a natural way to go with non-occurrent beliefs), and if we treat some of them as the result of what I call the "sublimation" of conscious content—the programming of the brain by the conscious mind.

So much for my arguments in favor of PGT. The rest of the dissertation is taken up with explaining how consciousness grounds content. My central proposal, introduced in **chapter 3**, is that phenomenal states ground intentional states when such states include phenomenal properties of a special sort: phenomenal-intentional properties (or "P-I properties"). P-I properties are experiential features of consciousness whose nature consists in the presentation to the subject of an intentional object. I argue that some of these have to be metaphysically primitive: there is no way to reduce them to non-intentional phenomenal properties.

I recognize the venerable distinction between *narrow* content and *wide* content. Narrow content is built up from P-I properties alone (or in connection with other phenomenal properties), whereas wide content is built up from P-I properties in connection with bits of the extra-mental world. Given this distinction, and given that a great deal of the intentional contents we entertain are wide, I only need to posit a *sparse set* of primitive P-I properties. P-I properties from this sparse set combine to form narrow modes of presentation of all the contents we can entertain. In **chapter 4** ("The Semantic

Structure of Phenomenal-Intentional States") I discuss the semantic combination of P-I properties: I show how *semantic* structure can be explained in terms of the *metaphysical* structure of phenomenal-intentional states. (I thereby address long-standing concerns about how monadicist theories of intentionality—those that treat intentional properties as monadic properties of the subject rather than as relations between the subject and something else—can accommodate semantic structure.) In **chapter 5** ("Abstracting and Constructing Contents") I discuss a number of types of intentional contents. I sketch how structures of P-I properties can serve as modes of presentation for intentional contents of these types, across perception, imagination and cognition.

In **chapter 6** ("The Emergence of Phenomenal Intentionality") I zoom out a bit. I have so far located intentional properties within the phenomenal domain; now I ask how we should locate phenomenal properties within a comprehensive metaphysics. I advocate for a version of ontological emergentism, on the grounds that (a) physicalism about phenomenal properties in general, and P-I properties in particular, doesn't look promising, and (b) panpsychism, emergentism's leading anti-physicalist rival, runs into problems related to the explanatory relevance of high-level facts, such as neural facts, to the dynamics of consciousness. (Like physicalism, it is too bottom-up). I advocate what I call "top-down property emergence": an emergent but nevertheless material subject instantiates conscious states, including phenomenal-intentional states, in response to brain-dynamics (as something like *interpretations* of the brain).

Acknowledgements

This project would not have been completed without the input and encouragement of many people. I cannot speak highly enough of Tim O'Connor, my dissertation chair. Tim has been available for consultation and encouragement (often moral as much as philosophical) whenever I have needed it. He has gently pointed out some rather embarrassing mistakes on my part, while insisting that my general line of inquiry remained interesting and important. The other three members of my committee—Colin Allen, Kirk Ludwig, and Fred Schmitt—have provided invaluable input as well.

Other faculty at Indiana University who have helped shape the project include Jonathan Weinberg (now at the University of Arizona) and Gary Ebbs. My dissertation's earliest recognizable ancestor was a paper I wrote for Jonathan's Philosophical Foundations of Cognitive Science course; Jonathan's apt criticisms of that paper, as well as Gary's skepticism about my inflationary intuitions more generally, illuminated just how question-begging my arguments were at the time. (This is not to say that Jonathan and Gary would now be fully dialectically satisfied with my arguments as I have developed them.)

I want to thank the members of the Metaphysicians' Dissertation Group at IU, whose continual goadings and guidings over the span of two years improved the quality of my thinking immensely. The MDG comprised Dave Fisher, Hao Hong, Tim Liesz, Nick Montgomery, Tim Perrine, and Harrison Waldo. Special thanks to Dave, my "logic consultant," whose readiness to devote time and energy to whatever problem had me stuck continually amazed me.

Philip Woodward

The Emergence of Mental Content: An Essay in the Metaphysics of Mind

Intentionality is the aboutness or directedness of mental states. According to the most popular theories of intentionality, a mental state's intentional content is constituted by its causal embeddedness in an organism, vis-à-vis that organism's environment. I argue that theories of this sort fail to explain how we could know the intentional contents of our mental states. As an alternative to causation-based theories of intentionality, I develop a *consciousness*-based theory of intentionality, as follows. *Phenomenal properties* are experiential aspects of consciousness. Among the various types of phenomenal property (sensory, somatic, conative, and so on) are *phenomenal-intentional properties*, or P-I properties. P-I properties are experiential aspects of consciousness whose natures consist in the presentation to the subject of an intentional content. In perception, imagination and cognition, P-I properties bind together to form modes of presentation of all of the intentional contents we can entertain. Along with the rest of the phenomenal domain, P-I properties emerge from the physical systems on which they depend, but are not reducible to, constituted by or realized in the states of those systems.

Table of Contents

Curriculum Vitae

1. The Acquaintance Argument for Intrinsic Intentionality

1. Introduction.

We creatures with minds are able to instantiate mental states that have intentional content.[1] Examples of intentional states include: believing that roses are red; contemplating the Pythagorean Theorem; intending to pick the children up from school; registering a stranger's gesture as a threat; imagining an ostrich; fearing one's mortality. For a state to be intentional is for it to be *about* or *directed at* something, such as a proposition, a property, an object, an event, or a state of affairs. Intentional directedness is one of the two traditional "marks of the mental" (the other being consciousness).

Intentional properties, as I will understand them, are those properties of intentional states that individuate such states by their contents, rather than according to their psychological modes (believing, wondering, hoping, imagining, thinking, etc.). Examples of intentional properties include: *meaning that p*; *representing F*; *being about/directed at o*, *referring to o*. The aim of this dissertation is to understand the nature of intentional properties—to understand, that is, how mental states can have intentional contents at all, and how mental states come to have the particular intentional contents that they have.

I will begin by focusing on intentional properties that are *purely qualitative* and *potentially conscious*. By *purely qualitative* intentional properties, I mean those that are not individuated in terms of concrete particulars. For example, the property of being

---

[1] I use the term "state" to refer to a property-instantiation at a time or to multiple property-instantiations at a time, by the thing in question. A *mental* state is the instantiation of *mental* properties by a thing at a time. The total mental state of a thing at a time is the cluster of mental properties instantiated by that thing; parts of this total state—proper subsets of the cluster of instantiated properties—are also states. I'm not picky about times: they can be momentary or not.

about unicorns and the property of being about stubbornness are purely qualitative intentional properties, whereas the property of being about *that* object, or being about *The Metropolitan Museum of Art* are not.

By *potentially conscious* intentional properties, I mean intentional properties that can be instantiated in a mental state for which there is something it is like to be in that state, for the subject whose state it is. Hence I am not directing my inquiry at those intentional properties instantiated in standing belief-states, non-accessed memories, suppressed desires, or representational states implicated in pre-conscious neural processing. It is of course a live possibility that the very same intentional properties are instantiated in conscious and non-conscious states.

Call such properties "paradigmatic intentional properties." Why focus my inquiry on intentional properties of just this sort? The reason is that they are the only intentional properties whose instantiation we are aware of from *within*, as follows. When, in contrast, we attribute intentional properties third-personally on the basis of an inference to the best explanation, the following questions can arise: Might a different, non-intentional explanation turn out to *better serve* the data? Might explanations in terms of intentional properties perform a practically indispensable but ultimately *merely heuristic* function? Might there be nothing more to *which* intentional properties are instantiated than the successful honoring of some *charity principle* or other? I contend that these challenges to realism do not and cannot come up for intentional properties whose instantiation we are aware of from within. To say that we are aware of instantiations of intentional properties from within is not to say that we have infallible introspective knowledge of the content of our mental states. It is to say that we could not be wrong about *whether* our mental states

have content, or about *whether there is a fact of the matter* about whether they have content.

My task in the present chapter is to argue for a constraint on any account of the nature of paradigmatic intentional properties. Such properties, I aim to show, are *intrinsic*—in a sense to be spelled out—to the conscious mental states that instantiate them. Before I argue for this constraint, I will discuss what I take to be the accounts of intentional properties that make use only of properties that have widespread traffic in the natural sciences (physics, chemistry, biology, etc.). It turns out that all of these accounts characterize intentional properties as non-intrinsic. Consequently, if my argument is successful, it follows that all of these accounts are false, at least when it comes to characterizing paradigmatic intentional properties. For simplicity, I will frame my discussion in terms of a generic intentional property, *being about F*, where F is a sub-propositional content (such as a property).

2. Naturalistic Theories of Intentional Properties

If you want a theory of intentional properties that invokes only concepts that enjoy good standing in current natural science, you need a strategy for "naturalizing" intentionality. This has been a central philosophical project in the philosophy of mind and cognitive science at least since the middle of the 20th century, and especially since the work of Fred Dretske, Jerry Fodor and Ruth Garrett Millikan in the 1980s. The basic idea has been to understand the intentional structure of the world as part of the causal structure of the world. We can identify four families of strategies that have emerged. All of these accounts characterize intentional properties as essentially *non-intrinsic* to the states in

which they are instantiated. If the arguments of the present paper are correct, it follows that all of these accounts are false, at least when it comes to characterizing paradigmatic intentional properties.

In order to understand what motivates these four theories, it is helpful to keep in mind a few of the features of intentional properties that any adequate theory has to account for. I'll mention three for now:

(1) *Specificity*. If a state has the content F, in general it's not the case that it also has—or is indeterminate whether it instead has—the content F-or-G; nor that it also has the content F*, where this is an extensionally equivalent content; nor the content F+, where this is a content that is the determinable of which F is a determinate; nor the content F-, where this is a content that is a determinate of which F is the determinable; nor the content {Fx, Fy, Fz...} where this is a proper subset of F-instances. A theory of intentional properties has a big strike against it if it cannot preserve these differences in content.

(2) *Availability*. Intentional properties do not appear and disappear in the mind lockstep with the environmental presence of their intentional objects. Rather, in imagination and cognition, contents can be brought to mind without any related environmental stimulus. A theory of intentional properties has a big strike against it if it cannot explain the ongoing instantiation of intentional properties despite arbitrary changes in the cognizer's environment.

(3) *Evaluability*. Intentional states are often evaluable for correctness or incorrectness. This is certainly true of perceptual states and belief-states. Consequently, intentional properties are such that their instantiation, when embedded in certain states on

certain occasions, partially constitutes *error* on the part of a cognizer. A theory of intentional properties has a big strike against it if it cannot explain how intentional properties can be instantiated in non-veridical intentional states.

Let's turn to the four naturalistic theories. (The importance of the three constraints just mentioned will crop up as we go.)

1. Causal covariance theories.[2] We can begin with the following two observations. First, we often exploit causal relations in order to transmit information about our environment and in order to extract information about our environment.[3] (Whenever we communicate via telephone we're so exploiting.) Second, we often attribute contents to cognizers on the basis of the causal relations she stands in with her environment. (Whenever we interpret someone's behavior in terms of her beliefs and desires with respect to her surroundings, we're so attributing.) The covariance theorist goes a step further and *identifies* intentional properties with certain causal relations that hold between a cognizer and the environment. Roughly, when some environmental feature covaries with the activation of a certain sort of structure within the cognitive system, that structure counts as a representing that property. We might think of the relevant type of activation as a state that plays the functional role of an existentially-quantified perceptual belief, e.g.: "Lo, a ___." (I will assume that a functional characterization of perceptual belief is forthcoming.)

So stated, the proposal does not honor the three adequacy constraints just mentioned. On the proposal, mental contents are *unspecific*: if the activation of a

---

[2] See e.g. Dretske (1981) and Fodor (1990).
[3] I'll be framing naturalistic theories in terms of causal relations, though some theorists (e.g. Dretske 1990) might prefer nomic relations. Pretty much everybody wants to reduce one to the other, so I don't think anything turns on my choice.

cognitive structure covaries with the instantiation of *being a fly* (to use a familiar example

from the naturalized-intentionality literature) it also covaries with the instantiation of

being a fly or a bb, and with much else besides, leaving that cognitive activation's

content under-specified. On the proposal, contents are *unavailable*: there is nothing in the

proposal that suggests how a content-bearing structure might be available for cognitive

use under *non-covarying* circumstances. And on the proposal, mental states are *non-evaluable*: there is nothing in the proposal that suggests how a content-bearing structure

could activate *erroneously*—how an activation could fail to have as its content whatever

environmental feature with which it *actually* covaries. So, to serve as a viable theory of

intentional properties, some additions will need to be made to our crude picture. Here's

how causal covariance theories, as I understand them, have been developed in order to

accommodate the desiderata: first, to accommodate Specificity, these theories include an

account of the *right sort of causal relation* that has to hold between a property-instance

and a mental state that represents that property instance. (For instance, the relation has to

be counterfactual-supporting; *mere* covariance isn't good enough.[4]) Second, to

accommodate Availability, the covariance relation needn't *actually* hold; it is enough that

the world be such that it *would* hold, were a property-instance to bear the relevant relation

to the cognizer.[5] Third, in order to accommodate Evaluability, a clause is added that

specifies the circumstances in which covariance ceases to fix content, because triggering

has happened erroneously. I'll call such circumstances an "E-circumstances." (Note that

---

[4] Says Fodor: that the covarying items covary has to be *a law of nature*.
[5] Why not: it is enough that it *has held*? That would be one way to go. I opt for a counterfactual rather than a historical construal of covariance theory (and the three other theories to be discussed presently) because I think that, so construed, naturalized theories of intentionality are in a stronger dialectical position with respect to the objection from semantic self-knowledge that I will be levying later this chapter. A cost, perhaps, of construing these theories in counterfactual terms is that it's possible for a system to have thoughts about some class of entities without having been in any causal contact with those entities.

if "E-circumstances" just are those circumstances ruled out by that causal relation invoked to satisfy Specificity, then Specificity and Evaluability will be accommodated in one go—a strategy that is attempted by some versions of causal covariance theory.)

So, here is the shape that a mature causal covariance theory takes: a frog's mental state represents *flyhood* if a structure included in that state would be activated in the (right sort of) causal presence of flies. Let 'O' stand for the relevant class of cognitive system;[6] let 'M_O'stand for a mental-state type that the relevant systems can instantiate; let '\$' stand for some content-bearing structure within a system;[7] let C stand for some causal relation; and let 'A' stand for the type of activation in 'O' that realizes the functional profile of perceptual belief:

> *Causal Covariance Theory* (CCT): M_O is about F if: \$ is part of M_O, and \$ would A were an F-instance to bear C to O (except in E-circumstances).

Activation A will get cashed out differently by different models of cognitive architecture. Perhaps the clearest picture is that supplied by the Computational Theory of Mind, on which A-ing is the tokening of \$ in the system's "perceptual-belief box". \$ itself could be a physical structure (a particular neuron, say, or the firing of that neuron) or something a little more abstract, such as a pattern of neural activation, or something *much more abstract*, such as a symbol in the language of thought (supposing that *being a symbol in the language of thought* can itself be naturalized).

Several accounts of E-circumstances can be found in the literature. According to Fodor, E-circumstances are those in which something other than an F-instance causes \$'s

---

[6] 'O' is for Organism, though I do not know what that category comes to; I assume for present purposes that there are objectively real things as cognitive systems. I return to the issue in chapter 2.

[7] Is \$ a type or a token? Theories will differ. Neurons are tokens, obviously. Mental symbols are types, presumably. In order to keep my characterizations as general as possible, I leave intentionally ambiguous whether '\$' should be read as picking out a token or a type.

A-ing—a G-instance, say—but would not have, were F-instances not apt to do so as well (i.e. G-caused A-ings are "asymmetrically dependent" on F-caused A-ings).[8] Dretske (1988) has suggested that E-circumstances are those in which G-instances cause $'s A-ing after the culmination of O's "learning period." Hill (ms) thinks we can just construct a list of E-circumstances.

Now, one can imagine an "open-question argument" for these proposals, as follows: Sure, G-caused A-ings are assymetrically dependent on F-caused A-ings; sure, in the learning period A-ings were only F-caused; sure, G-caused A-ings are among those mentioned in the list of deviant cases. But whence the *incorrectness* of G-caused A-ings? Where does the semantic normativity come from? I'm not sure whether the covariance theorist can have anything informative to say at this point. Perhaps she ought simply to deny that the relevant normative question is open.[9]

2. System-role theories.[10] The theories I have in mind are versions, or at least close cousins, of covariance theories, in that they identify intentional properties in terms of causal relations that hold between states of a system, on the one hand, and features in the system's environment, on the other. But they deal with Evaluability otherwise than by adding a clause specifying E-circumstances. What makes it the case that a system can misrepresent—that its G-caused A-ings don't automatically endow $ with a disjunctive content F v G—is a matter of the system's itself having certain representational *purposes*. $, that is, has the *function* of representing F.

---

[8] I discuss asymmetric dependence at some length in the chapter 2.
[9] Bickhard (2004) takes the lack of an informative response to the open-question argument as a serious liability for covariance theories.
[10] See e.g. Dretske (1986) and (1995); Bickhard (2004) and (ms).

As best I can tell, two strategies have been proposed in the literature for grounding the relevant teleological properties: one built around a system's *homeostatic needs*, and another around a system's *evolutionary ancestry* (i.e. the selection pressures that caused ancestor-systems to evolve in the way they did). System-role theorists take the first strategy, which is our present concern; and I'll discuss the second strategy in a moment. According to system-role theories, a frog's mental state represents *flyhood* if part of that mental state is such that the frog *needs* it to be activated in the causal presence of flies. More precisely,

> *System-Role Theory (SRT)*: M_O is about F if: $ is part of M_O, and it would contribute to O's homeostasis were $ to A in the event that an F-instance bore C to O, and it would detract from O's homeostasis were $ to A in the event that a non-F-instance bore C to O.

System role theories thus tie intentionality to usefulness; a state of an organism has a content when that state guides the organism toward survival-conducive functioning.

Dretske illustrates a version of the view by way of a discussion of marine bacteria that possess internal magnets. Like compass needles, these magnets are responsive to the earth's magnetic field:

> Since these magnetic lines incline downwards (towards geomagnetic north), bacteria in the northern hemisphere (upwards in the southern hemisphere), oriented by their internal magnetosomes, propel themselves towards geomagnetic north. The survival value of magnetotaxis (as this sensory mechanism is called) is not obvious, but it reasonable to suppose that it functions so as to enable the bacteria to avoid surface water. Since these organisms are capable of living only in the absence of oxygen, movement towards diamagnetic north will take the bacteria way from oxygen-rich surface water and towards the comparatively oxygen-free sediment at the bottom...If a bar magnet oriented in the opposite direction to the earth's magnetic field is held near these bacteria, they can be lured

into a deadly environment...this appears to be a plausible instance of misrepresentation.[11]

The thought here is that a type of activation of a bacterium's magnetostatic system counts as *representing geomagnetic north*, even when that activation is caused by a stimulus other than geomagnetic north. Its so counting is a matter of that activation's being implicated in goings-on in the bacterium that are helpful to it only when that activation *is* caused by geomagnetic north. So, given its homeostatic needs, it would be *wrong*, semantically, for it to similarly activate its magnetostatic system when caused by a bar magnet—a stimulus that might be proximally equivalent (save for the direction whence it stimulates) as that of geomagnetic north.

Even in the context of Dretske's primitive example, some pretty obvious problems emerge. First of all, there is a direct connection between (a) a bacterium's receiving a causal stimulus from the earth's magnetic field, (b) the activation of its magnetostatic system, and (c) its steering clear of a deadly threat (oxygen). But what about activations within the bacterium, caused by other aspects of its environment, that have neutral or at any rate less dire consequences for its behavior? Whether an activation in the bacterium *contributes to* or *detracts from* its homeostasis may be a matter of degree, and for some environmentally-triggered activation-patterns there may be no stable fact of the matter as to whether they help or hurt. In these cases, it is not clear that SRT has the resources to accommodate Evaluability.

Further, whether an activation-pattern helps or hurts will depend on the nature of the bacterium's environment. As Dretske points out, a magnetosome will lead a native northern-hemisphere-dwelling bacterium to an untimely death if that bacterium is re-

---

[11] Dretske (1986), p. 164.

located to the southern hemisphere. In that environment, the bacterium's homeostatic needs vis-a-vis responding to stimuli from the earth's magnetic field are reversed. While it's helpful, in the northern hemisphere, for a bacterium to have its magnetosome activate in response to stimulus from geomagnetic north, it is decidedly unhelpful for that same activation to take place in the southern hemisphere. Here's what *would* be helpful: when in the northern hemisphere, the bacterium "takes" (so to speak) its magnetosome's activation-pattern to indicate the absence of oxygen, and when in the southern hemisphere, to indicate the *presence* of oxygen. That is to say that it would serve the organism's needs for the same activation-pattern to have different representational contents*, depending on the environment*. But how are environments to be individuated? Does the presence of a malicious bar-magnet-waver make for a different environment? If so, it looks as though we were wrong to say that the bacterium misrepresents geomagnetic properties, when in the presence of a bar-magnet; rather, the bacterium has entered a new environment, with respect to which its homeostatic needs have changed. In this new environment it *correctly* represents the presence of bar-magnetic properties. (It then proceeds to act stupidly on the basis of its correct representations, rather than rationally on the basis of incorrect ones.) In sum: an organism has homeostatic needs only relative to an environment. If SRT is going to be viable at all, it will need a nontrivial way of distinguishing an organism's normal or natural environment from non-normal environments, and then it will need to index contentfulness to needs-in-normal-environments.[12]

---

[12] One strategy (which Dretske discusses) is to understand an organism's normal environment as the environment of its phylogenetic development. The resulting view will be similar to—perhaps a version of—adaptive role theories.

Another concern is that it is not clear how an organism's pragmatic concerns could carve contents finely enough to satisfy Specificity. A bacterium in a certain environment needs to be causally sensitive to geomagnetic forces so as to avoid deadly oxygen. Perhaps it is right to say that it does *not* thereby need to be sensitive to geomagnetism-or-horsehood, and to this extent SRT is *somewhat* specific. But it will be just as good for the bacterium to be sensitive to geomagnetism-during-its-lifetime. Worse, it would be *better* for the bacterium to be sensitive only to geomagnetism-that-actually-correlates-with-the-presence-of-oxygen. The system-role theorist owes us an account of the pragmatic downsides of representing such awkwardly broad or narrow categories.

3. Adaptive-role theories. As we have seen, a natural way to respect Evaluability is to appeal to a system's representational *purposes*: to say that a system commits semantic errors is to say that it fails in these purposes. One way to naturalize such talk is to understand it in terms of the needs of an organism. Another is to understand it in terms of the metaphorical *design* of the organism, i.e., in terms of its evolutionary history. According to such a view, a frog's mental state represents *flyhood* if the frog's ancestors gained a selection advantage by having a part of type-identical mental states properly activate in the (right sort of) causal presence of flies. More precisely,

> *Adaptive-Role Theory (ART)*: M_O is about F if: $ is part of M_O, and it contributed to O's ancestor's selective advantage when ancestral correlates of $ A'ed in the event that F-instances bore C to O's ancestors, and it detracted from O's ancestor's selective advantage when ancestral correlates of $ A'ed in the event that non-F-instances bore C to O's ancestors.

A striking feature of adaptive-role theory is that it ascribes content only to systems that have evolutionary ancestors; artificial or accidental systems could not, according to a straightforward reading of such theories, instantiate intentional properties, or at least the same intentional properties natural systems instantiate. Those who think that philosophical theories of intentionality should be as beholden as possible to scientific usage of intentional notions might find this peculiarity unproblematic (since all uncontroversial examples of representational systems *do* have evolutionary ancestors). But if non-evolved cognitive systems are possible, and if a metaphysics of intentional properties needs to explain possible as well as actual instances, then it will be hard to see how adaptiver-role theory could provide a fully general account of intentionality.

A related worry has to do with adaptationism generally. The process of natural selection is messy. Some biological structures have evolved via adaptation: they have been subject to gradual calibration as a result of a selective feedback-loop. But there are other, non-adaptive evolutionary processes. "Exapted" structures are such that their present-day contribution to a species' viability has little to do with the selection-pressures on the adapted structures from which they descended. Some structures (so-called "spandrels") are the bi-products rather than the descendants of adapted structures. In short, adaptive-role theory ascribes intentional properties to organisms only insofar as those organisms have representational purposes, and they have representational purposes only insofar as their cognitive mechanisms were adaptively calibrated—an empirically open matter.

Finally, adaptive-role theory might run into the similar troubles with Specificity as did SRT. The question of how to specify the content of an organism's mental state will

turn on the question of how to specify the causal forces that guide natural selection.

Suppose that we want to explain a frog's fly-catching in terms of its representational abilities. According to ART, we should appeal to the selective forces operative on the frog's ancestors. It is natural to say that the frog's ancestor's survived *because of their sensitivity to the presence of flies*. But suppose an alternate explanation were proposed in terms of sensitivity to the presence of flies-in-the-neighborhood, or in terms of flies-or-comets. In what sense is the more natural-sounding explanation *correct*?[13] Here is one sense: the more natural explanation is more parsimonious, elegant, convenient, etc. But if whatever contributes to or detracts from selective success is partly a matter of out theoretical interests, then, on ART, which intentional properties a system instantiates are partly a matter of our theoretical interests. And so saying would require the abandonment of a realist position—an unacceptable move with respect to paradigmatic intentional properties. Here's a second sense (in which the more natural-sounding explanation could be the *correct* one): the property it invokes is metaphysically more natural.[14] This suggestion has the ring of sound objectivity to it, but I'm not sure that it is correct. Note: adaptative-role theories do not appeal to counterfactual relations between organisms and their environments, but rather to *actual* causal relations that have held between organisms and their environments. The *actual, historical selection pressures* on paleo-frog survival might most accurately be represented by a list of precisely characterized events—*not* by a natural category.[15] At any rate, in order to satisfy Specificity, a viable version of ART

---

[13] See Millikan (1990), p. 334, for a discussion of this issue.
[14] This idea comes from Lewis (1983).
[15] Christopher Peacocke (1992, p. 130) raises a similar worry, which he calls "the problem of reduced content": "What explains the proliferation and survival of the belief-producing mechanisms and the organisms containing them when p is believed, p is true, and all is working properly is the truth of all of the (logical) consequences of p that have a causal impact on the organism....how is the teleological theorist to block an incorrect assignment of content to beliefs, namely one that requires for its truth merely the truth of

will require a principled way to sort the natural, more general sorts of adaptational explanations from the unwieldy, idiosyncratic ones.

    4. Conceptual-role theories. The three families of theories we have just looked at all pin down content in terms of property-instances instantiated in the environment: how an organism would respond if stimulated by those property-instances (CCT), how an organism needs to respond if stimulated by those property-instances (SRT), and how an organism's ancestors responded when stimulated by those property-instances (ART). Conceptual role theories set aside this strategy entirely. Instead, they pin down content in terms of intra-systemic relations. According to these theories, a frog's mental state represents *flyhood* if that state's functional relations with other mental states map onto the inferential relations that hold between various fly-related propositions. It will be helpful to change the notation slightly. Let 'B_1', B_2' and 'B_3' stand for types of activation that play the functional role of belief in O, and let '$B(\$)$' (and suchlike) stand for an activation-type that implicates $\$$ so as to functionally realize a belief about or involving whatever $\$$ denotes. Let 'P_1', 'P_2' and 'P_3' stand for propositions, and let '$P(F)$' (and suchlike) stand for propositions about or involving Fs or F-ness.

> *Conceptual-Role Theory (CRT)*: M_O is about F if: $\$$ is part of M_O, and were C-relations to hold between $B_1(\$)$, $B_2(\$)$, $B_3(\$)$, etc., these relations would be isomorphic with I-relations between $P_1(F)$, $P_2(F)$, $P_3(F)$, etc. (except in E-circumstances.)

Precisely because CRT does not make mention of environmental states of affairs, it is uniquely suited to contents that do not correspond to physical properties, e.g. the identity-

---

all the logical consequences of p that have a causal impact on the thinker, rather than the stronger condition of the truth of p itself?"

relation or the addition-function. So, for example, according to CRT, a state's having the content *plus* is a matter of certain counterfactuals' being true of it, viz.: were it functionally connected in the right sort of way with states with the numerical contents *m* and *n*, it would causally contribute to the activation of a state whose content is the sum of *m* and *n*.

On the other hand, CRT faces especially formidable obstacles to satisfying Evaluability and Specificity. For the three previously discussed theories, satisfying Evaluability is a matter of distinguishing between those environmental-organismic causal relations that are content-fixing vs. those that are error-conducive. Any perceptual belief generated by a content-fixing causal chain is guaranteed to be *true*, in other words. But matters are more complicated in the context of CRT. Content-fixing causal relations are not environmental-organismic connections, but rather connections between states *within* the organism. And there are two distinct ways for these transitions to fail to be content-fixing. First, they can correspond to *empirical* rather than *conceptual* inferences. For example, if I believe that flies are carriers of disease, then my fly-ish thoughts might be such that they frequently generate disease-carrier-ish thoughts; but it's odd to say that my fly-concept should be *analyzed in terms of* disease-carrying, and in terms of every other feature I associate with flies. Thus Fodor & Lepore (1991) argue that CRT thus faces a dilemma: either CRT implies that meaning is implausibly holistic, according to which every belief an organism has about a subject-matter (save for beliefs with existential implications) is analytically true, or else CRT requires that there be a distinction between empirical and conceptual inferences, a distinction couched solely in causal terms.[16] Now,

---

[16] Fodor & Lepore note that it is hard enough to couch such a distinction in *semantic* terms—if the last fifty years' attempts at establishing an analytic/synthetic distinction are any indication.

I am not sure but that Fodor's own work provides the CRT-advocate with just the tools to distinguish between empirical and conceptual inferences in causal terms, viz. by reference to the asymmetric dependence of the former on the latter. For example, it is plausible that were my fly-ish thoughts not to generate insect-ish thoughts, they would not generate disease-carrier-ish thoughts, but were fly-ish thoughts not to generate disease-carrier-ish thoughts, they still would generate insect-ish thoughts. I am uncertain whether an inference's asymmetric dependent on another is a sufficient condition for its being a merely empirical inference; perhaps some conceptual inferences are asymmetrically dependent on other conceptual inferences. At any rate, the strategy seems worth exploring.

But there is another way for a transition to fail to be content-fixing: it can fail to be *rational*. Irrationality is a source of negative semantic assessment that does not come up for the previous three theories, since those theories do not involve *inferences*: perceptual beliefs can be false, but they cannot be irrational. Notice how crucial it is that only rational transitions be content-fixing. So long as it is an open matter whether a content-fixing transition is rational or irrational, is utterly indeterminate what content each state expresses, no matter how large or complex the system. To see this, just think of the system as committing really stupid non-sequitors over and over again. Since every proposition bears the relation *being a consequence of or not being a consequence of* to every other proposition, an irrational system's transitions can be mapped onto transitions between any propositions whatever. Notice, at the same time, how difficult it is to sort a system's rational vs. irrational transitions in purely causal terms. Rationality is an irreducibly semantic property of transitions between intentional states. What makes it the

17

case, that is, that one state rationally follows from another has to do with the content expressed by those states. Content is metaphysically prior to rationality; but it seems that CRT needs the reverse to be the case in order to get off the ground. (Note that asymmetric dependence definitely won't help here: empirical inferences can be asymmetrically dependent on conceptual inferences; they are not thereby irrational. So rational inferences can be asymmetrically dependent on other rational inferences.) I see no way forward here. There is an aspect of Evaluability that CRT simply lacks the resources to satisfy.

Specificity also poses a more complicated challenge to CRT than to the three theories previously discussed. For those theories, satisfying Specificity is a matter of distinguishing which candidate feature, of the many that play a role in any particular content-fixing causal chain, serves as the unique semantic content of the structure that's at the organismic end of the chain. For any particular causal chain, not just any feature gets to be a candidate, of course—many features clearly play no role whatsoever in any given causal chain and can thereby be ruled out. Not so with CRT, since a content-bearing structure need not be causally connected to (an instance of) what it denotes. Now, if most contents were such that (a) in order to think about them, subjects had to know a lot of conceptual truths about them, and (b) the inferential relations among conceptual truths about a content were sufficient to distinguish that content from others, then the lack (on CRT) of a causal connection between content and content-bearer might not be a problem. But (b) is implausible and (a) is demonstrably false: many contents are such that we need know *hardly any* conceptual truths about them. At any rate, natural kinds—such

as *fly*—seem to be such that *just about everything we believe* about them is subject to revision.[17]

The lesson, I think, is that CRT cannot stand on its own as a naturalized theory of intentionality. But that does not mean it should be dismissed. (If it did, naturalized intentionality would be in a bad spot seeing as none of the other theories supply the resources to ascribe any contents besides physical features.) The obvious alternative is to combine CRT with another theory. According to this strategy, *some* of a system's content-bearing structures have the content they do because of the way they causally depend on other states; but some of these other states have content *already*, as specified by one of the other theories.

It would not be fruitful here to explore all the possible ways that the four theories could be combined, to construct a comprehensive theory of content. I will briefly sketch a hybrid version that looks to me as strong a version as any. Here is the basic idea: a cognitive system has representational *modules* with particular representational purposes, where these purposes are fixed either by way of evolutionary history or homeostatic needs. The purposes of the modules will be somewhat general, e.g. *to indicate objects*, *to indicate causes*, *to make conceptual inferences*, etc. These modules will generate particular representations, whose content is not fixed by the general representational purposes of each module, but rather by (a) patterns of causal dependence between representations and properties in the environment (per CCT), or (b) patterns of causal dependence among representations (per CRT). Presumably these meager representational resources can be combined to form new representations and to encode new

---

[17] So Kripke (1980) convincingly argues.

representational purposes, on the basis of which an intentional structure of arbitrary complexity and power can be iteratively built.[18]

The aforementioned four theories comprise the strategies for naturalizing intentional phenomena that have been pursued over the last thirty years or so. My purpose has not merely been to review the literature, but to map the relevant conceptual space. That is, there is good reason to think that any naturalistic theory of intentionality is going to amount to a version of one of these four and/or a combination thereof. Consider the resources these theories draw on: patterns of causal dependence that hold among (a) internal states of a system; (b) internal states of system and features of the external world; (c) current states of a system, states of ancestral systems, and features of the external world. If the goal is to locate intentional phenomena within the ambit of causal phenomena (actual and counterfactual), it does not look as though there are any raw materials that the naturalizers have simply skipped over. No, the materials are all out on the table; remaining philosophical work will amount to refinement and assembly.

Our exposition of naturalized intentionality is almost complete, but not quite. We have framed the four theories in terms of *sufficient conditions for a mental state's having some intentional property*. But intentionality has not been "naturalized"—shown to be locatable within the ontology of the natural sciences—until the *natures* and not just the *instantiation-conditions* of intentional properties have been framed naturalistically. That is, it is consistent with the characterizations so far given of the four theories that intentional properties are non-natural properties that supervene on the conditions specified in one or more of the right-hand side of the conditionals specified above. So, how can we extract a naturalistic metaphysics of intentionality from the characterizations

---

[18] Carey (2009, ch. 13) develops a composite theory along these lines.

given in the present section? Here is one way: *identify* intentional properties with the functional/dispositional properties delineated by the right hand side of one or more of the schemas above. Going this route would have the odd result that *there is no property* of primitively being about something; that is, not even God can think about F-ness without instantiating the right sort of functional/dispositional property. A second approach would be to say that the relevant functional/dispositional properties merely *constitute* intentional properties, thus amounting to one of multiple ways that intentional properties can be "realized." But absent a clear understanding of the realization-relation, such talk is not particularly illuminating. I propose a third option: advocates of naturalized intentionality should understand their views as characterizing a *species* of intentionality, i.e. one way among multiple ways (the set of which comprise a strongly unified family of properties) for an item to be about another item. Call intentional properties of this species "naturalistic-intentional properties", or "N-I properties" for short. To advocate for one or more naturalistic theories is to contend that the intentional properties instantiated by *us* (or whatever cognitive systems amount to the relevant theoretical target) are N-I properties— i.e. one or the other, or a combination, of the following four types (characterized more schematically than above):

(1) *N-I(CC) property* = the presence (in a state of a O) of $ such that $ would A were O causally stimulated in certain ways.

(2) *N-I(SR) property* = the presence (in a state of a O) of $ such that O needs $ to A were O causally stimulated in certain ways.

(3) *N-I(AR) property* = the presence (in a state of a O) of $ such that O's ancestors needed their $-correlates to A when they were causally stimulated in certain ways.

(4) *N-I(CR) property* = the presence (in a state of a O) of $ such that $B_1(\$)$, $B_2(\$)$, $B_3(\$)$, etc. would cause each other's activation in O in certain circumstances.

3. The Acquaintance Argument

In the remainder of the chapter I argue that paradigmatic intentional properties are not N-I properties. I lay out my argument briefly here and then proceed to argue at length for its premises.

(1) Some paradigmatic intentional properties are such that subjects can be introspectively acquainted with them.

(2) Subjects can be introspectively acquainted only with the intrinsic features of their conscious mental states.

(3) N-I properties are not intrinsic features of conscious mental states.

(4) Hence, some paradigmatic intentional properties are not N-I properties.

Call this "The Acquaintance Argument." I will address its premises in reverse order—from least controversial to most controversial.

Premise 3: N-I properties are not intrinsic features of conscious mental states. For my purposes, a property is intrinsic to a conscious mental state only if it supervenes on the essential properties of that state—i.e., the properties that make it the very type of conscious mental state that it is. So the question at issue is whether N-I properties supervene on the essential properties of conscious mental states. Here is one way we

might answer the question: No—the essential properties of conscious mental states are *phenomenal* properties, and there is no necessary connection between phenomenal properties and the functional/dispositional properties relevant to N-I properties. Whatever the virtues of this response, it isn't dialectically appropriate here, because what drives it is the epistemic gap that holds between phenomenal states and functional states *generally*.

In order to evaluate the question in a dialectically appropriate way, we will need to begin not from a phenomenological conception of conscious mental states but from a functional conception of conscious mental states.[19] There is disagreement over how best to do this, but I gather that at least three components will be involved: (1) a functional specification of the content of conscious mental states (which we have at hand in the form of N-I properties); (2) a functional specification of psychological modes; and (3) a functional specification of a state's *being conscious*. Likely (3) will be framed in terms of a state's *occurrency*—i.e., what makes it the case that the state of a system is *activated* in a special way. So: suppose a structure $ is part of an occurrent state, where $ has the content *being a snake*, per one of the four naturalized theories; and that the occurrent state as a whole amounts to a conscious fear that one will see snakes at the zoo. Our question is whether its N-I properties supervene on it essential features.

It is immediately clear that any N-I(AR) properties it has do not so supervene. Such properties partly supervene on the *actual phylogenetic histories* of systems. A system's tokening some particular occurrent state just is not a matter of the system's having a certain evolutionary history. Matters are not so obvious regarding the other three types of N-I property, since those properties are not defined in terms of actual causal

---

[19] It is open to the functionalist to maintain that the two conceptions converge on the very same type of state, despite our being unable (at present, or maybe ever) to connect the two conceptions in an explanatorily perspicuous way.

histories, but rather in terms of dispositions of $ to be activated by the system under

certain conditions.[20] Nevertheless, I think it is evident upon reflection that N-I properties

of the other three sorts also fail to supervene on the essential features of states that

instantiate them. Here's the basic idea: there is no necessary connection between $'s

being tokened in an occurrent state, on the one hand, and its playing the content-fixing

role that it plays in the system, on the other; that it plays its content-fixing role is

grounded in more than the occurrent state itself. For example: according to CCT, an

occurrent state is about snakes if it contains a structure that would be tokened as part of a

perceptual belief in the event that the system were to stand in the right sort of causal

relation to snakes. But what makes it the case that $ has precisely that

functional/dispositional profile turns on the system's *having the right sort of sensory*

apparatus in the right sort of environment. If $ were embedded in a very different system

and/or a very different environment, it would *not* be disposed to be tokened within a

perceptual belief in the causal presence of snakes. Similar considerations apply when it

comes to SRT (which brings in wider organismic needs as part of the supervenience-base

for N-I properties) and CRT (which brings in rational transitions between occurrent

states). In short: N-I properties are not intrinsic to the occurrent states that token them,

---

[20] Any version of naturalized intentionality that *does* appeal to actual causal histories has the immediate
implication that intentional properties are never intrinsic (thus falling prey to the Acquaintance Argument).
Though naturalistic theories of intentionality are sometimes characterized in terms of causal histories
(owing, I gather, to their resemblance to causal theories of reference), I have not characterized them that
way, in part to avoid this immediate implication.
      Here are two other reasons that one might give (but that I reject) for counting N-I properties as
extrinsic: (1) N-I properties are dispositional, and dispositions never locally supervene, but instead are
instantiated only relative to the laws of nature. (2) N-I properties are dispositional, but only categorical
properties are instantiated in conscious mental states. I reject these lines of argument because I accept that
there are primitively dispositional properties and because I think that some dispositions are phenomenally
given (for example: disposing me to scratch is part of what itches essentially feel like). Advocates of
naturalized intentionality who reduce dispositions to laws and who deny that phenomenal properties are
essentially dispositional will feel the pull of (1) and (2).

because the functional/dispositional profiles individuative of N-I properties are fixed by broader, non-occurrent features of the system.

I can think of two ways that a defender of the intrinsicality of N-I properties might push back. First, she might say that since a mental state's intentional content is essential to that state, the supervenience-base for its intentional properties must be essential to that state; and hence all broader systemic/environmental states of affairs that form part of the supervenience-base *are*, after all, somehow constitutive of the state. One's conscious fear that one will see snakes at the zoo turns out to be metaphysically dependent on one's having the right sort of sensory apparatus, being suitably embodied, having a cognitive system with right sort of inferential capacities, etc. In short, tokening an occurrent state implicates much more of the world than we may have thought it did.

In reply: it is one thing to say that some state's occurring entails much else besides (with nomological, metaphysical, or even logical necessity). It is another thing to say that all of these entailments are *essential to* the state. For example, the lifting of my arm at noon on Tuesday might entail both that I have an arm at noon on Tuesday and that atmospheric oxygen/carbon dioxide levels are within a certain range at noon on Tuesday; plausibly, the former but not the latter is essential, in my sense, to the event in question. Similarly: for all I know, I could never have acquired the concept of a snake were I not suitably hooked up to a snake-saturated environment; nevertheless such causal embeddedness is not *what it is* to have a conscious fear of snakes.[21] Now, I am not sure

---

[21] So-called "wide" mental content is content that does not supervene on intrinsic states of an organism, but instead supervenes on bits of the world as well. By denying that anything external to an occurrent state is essential to that state, am I committed to denying that there is wide mental content? Not at all. What I am committed to denying is that wide content is *essential* to the mental states that have it. And this an independently plausible thing to say. Whether Oscar's mental state picks out H_2O or XYZ is a contingent matter—a matter of which environment he has grown up in.

what criterion a functionalist ought to apply in order to distinguish between what is and is not essential to an occurrent state. Note, though, that some such distinction has to be made on pain of preserving the occurrent/latent distinction. An occurrent state of a system is *what the system is doing*, not *what the system is structured to be able to do*— even if what it is doing is deeply dependent on its having a certain structure.

A second way of defending the intrinsicality of N-I properties would be to alter the way that N-I properties are characterized. For example: instead of saying merely that an N-I(CC) property is the presence (in a state of a O) of $ such that $ would A were O causally stimulated in certain ways, one could add: *were $ embedded in a system of such-and-such a sort and were O embedded in an environment of such-and-such a sort.* (We might say that the strategy works by attributing to $ a *context-relative intrinsic disposition*.) The idea here would be to render $'s content-fixing functional/dispositional profile necessary to it, not by expanding the class of items essential to states that include it (per the previous defensive strategy) but by precisifying its functional/dispositional profile so that it has that profile *regardless* of which sort of system embeds it.

The strategy feels like a cheat: if it works, then I have the *instrinsic* property of being disposed to be-thought-highly-of-by-the the-President-were-the-President-to-think-highly-of-me. Whether or not someone would think highly of me is surely not intrinsic to me! But there is a deeper problem with the strategy. If a structure has *one* context-relative disposition, it has *countless* context-relative dispositions. That is, in some contexts $ would be activated by snakes; in others, by frogs; and so on forever. So it will turn out that a state that includes $ will have *all of these contents*. That's the cost of making it such that $ has its context-fixing functional/dispositional profile necessarily.

I conclude that N-I properties are not features of the conscious mental states that instantiate them.

Premise 2: Subjects can be introspectively acquainted only with the intrinsic features of their conscious mental states. Introspective acquaintance is that direct cognitive relation that a subject bears to her conscious mental states when, and because, she attends to them.[22] Acquaintance, as I am understanding it, is not reducible to any propositional attitude or to any ability, but is rather a form of thing-knowledge, of *knowing-what*. Acquaintance is a species of cognitive relation, of which there are many (imagining, sensing, thinking about, and so on), but it is unique among cognitive relations in that its cognitive object must actually exist. If S thinks about x, it does not follow that x exists. If S is acquainted with x, it *does* follow that x exists.[23]

While Premise 2 strikes me as obvious, I doubt it will strike everyone that way. But I think it can be derived from some facts about introspection that are widely accepted—facts which shed light on why it seems so strange to deny it. It is well known that one's judgments about one's own mind are not infallible.[24] There is, nevertheless, a certain form of fallibility that a certain type of introspective judgment cannot have: when one forms an introspective judgment on the basis of an act of introspective acquaintance

---

[22] I will sometimes drop the modifier 'introspective'. As a matter of fact I do not think there are any other types of acquaintance (unlike e.g. Russell, who thinks we can be acquainted with universals).

[23] This is because the cognitive relation of introspective acquaintance is always underwritten by a *metaphysical* relation, which we might neutrally characterize as presence-to-the-subject. That is, S can be acquainted with x only if x is phenomenally present to S. According to Russell (1910), the acquaintance relation just is the presentation-relation in reverse. Because I connect acquaintance to *attention*, and because I think subjects are presented with more than they attend, I take presentation to be a necessary but not sufficient for acquaintance.

[24] There are many sources of fallibility. One can self-attribute motivations for one's behaviors, where such self-attribution is more or less confabulatory, a game of self-interpretation. One can attribute to oneself the belief that p on the basis of one's judgment that p, despite the fact that one lacks a sufficiently stable credence level with respect to p. One can fail to attend sufficiently carefully to the state one proceeds to report. One can lack concepts of sufficient fine-grainedness to accurately capture one's mental states. One can forget what one has just attended, during the time-lag between introspection and the forming of an introspective judgment.

27

with a conscious mental state, one cannot go wrong *by being misled by the appearances*. In introspective acquaintance, appearance and reality go together. Hence acquaintance-based introspective judgments are not subject to what Terry Horgan (2012) calls "appearance/reality fallibility".

By contrast, judgments about non-intrinsic features of our conscious mental states *do* admit of appearance/reality fallibility. This is because non-intrinsic features of a mental state supervene on states of affairs external to that state, i.e. states of affairs that are no more closely connected to that state than via a causal connection. And appearances can mislead, when it comes to judgments about items that are no-more-closely-than-causally connected to one's mental states: the familiar effects of a familiar cause can be replicated by a *distinct*, non-familiar cause (even if it would take an evil demon to pull off the stunt.) It follows that one can only form acquaintance-based introspective judgments about items that are more-closely-than-causally-connected to one's mental states—i.e. their *intrinsic* features, those features that supervene on nothing outside of the states in which they are instantiated.

Premise 1: Some paradigmatic intentional properties are such that subjects can be introspectively acquainted with them. I defend Premise 1 via the following sub-argument:

(a) It is possible for a subject to cross-modally recognitionally sort her mental states with respect to their intentional content.

(b) The only adequate epistemic explanation of such a capacity is in terms of a subject's introspective acquaintance with her intentional states.[25]

---

[25] By "epistemic explanation" I mean an explanation of the epistemic status of a cognitive state.

(c) Such a capacity calls out for epistemic explanation (it is neither epistemically trivial nor epistemically brute).

Here's an illustration of the capacity invoked in (a). A boy works as an apprentice to a potion-master. The potion-master makes concoctions of sundry ingredients, samples of which he has meticulously collected over his many years. The demands on his time prevent him from ever cataloguing his collection. Instead, he stuffs his findings in whatever receptacles he can get his hands on and then piles the receptacles in the cupboards in his pantry, never forgetting which ingredients are in which receptacles and where they are located in the pantry. While the potion-master is at work, he calls out identifying descriptions of the receptacles to his apprentice. "Bring me the glass vial on the third shelf with the blue liquid in it! Bring me the wooden crate in the second cabinet with the black beetles in it!" The apprentice dutifully runs to the pantry, retrieves the items that satisfy his master's descriptions, and delivers them.

My focus is on the apprentice's cognitive process that allows him to succeed at the task. First, he has an auditory experience as of the potion-master's vocalizations, and understands those vocalizations as linguistically encoding a certain content.[26] Next, when he enters the pantry, he has visual experiences as of the cupboards and all they contain. Finally, he visually identifies those items that satisfy the linguistic content of his master's vocalizations. That is: when he finds the item his master sent him into the pantry to find, he reidentifies the intentional contents of his visual experience as of the same sort as the intentional contents of his auditory experience or, at any rate, of his conscious understanding of the linguistic content of that experience. He is able to tell that one of his

---

[26] I remain neutral with respect to whether linguistic content is part of the apprentice's auditory experience, is a distinct but concomitant experience, or is inferred by him (however rapidly and automatically) on the basis of his auditory experience.

experiential states shares intentional properties with another of his experiential states: he can *cross-modally recognitionally sort* conscious mental states according to their intentional properties.

One might respond: "The apprentice doesn't compare his visual experience to the Master's descriptions; he compares *the items in the pantry* to the Masters descriptions." He does so, yes, *by way of* recognitionally sorting his experiences. (Compare: he *sees* the items in the pantry by *visually representing* them.) Here are two reasons to think the two epistemic tasks (sorting experiences, sorting experienced items) are not in competition. (1) Phenomenal experiences present worldly items, but also present themselves. The very same attentional act can therefore count as perceptual and introspective.[27] (2) Were the apprentice hallucinating the whole affair, he could still perform an epistemic task of the same type.

Why think that (b) is true? Why think, that is, that the right way to characterize the apprentice's success is in terms of his being introspective acquainted with intentional properties? The basic thought here is that if the apprentice's recognitional capacities are in any sense *epistemically guided*, then he must be acquainted with the items mentioned in those judgments, because only states with which he is acquainted could do the guiding. Now, it is not in general the case that if a subject can sort two items X and Y for sameness or difference, then the subject is acquainted with X and Y. We are sometimes able to "mindlessly" or distractedly sort two items: a bored assembly-line worker might successfully discriminate machine parts in such a manner; an expert pianist might discriminate piano keys in such a manner. But I don't think such capacities call for epistemic explanations; their reliability is to be explained in terms of the subject's

---

[27] For interesting discussion, see Nida-Rumelin (2011) and Siewert (2012).

execution of a skill, *not* in terms of the subject's performance of an inference-like rational transition.

By way of illustrating this distinction, consider our ability to perceptually recognize faces. For most of us, facial recognition is totally automatic—it is as obvious to us *whose* (familiar) face we are looking at as it is *that we are looking at a face*, and if we were pressed to justify our recognitional judgment, we might have nothing to say. I gather that our capacity for facial recognition need not be *guided*. For prosopagnosics, by contrast, facial recognition is an epistemically substantive capacity. Their perceptual experiences of faces do not, by themselves, represent *whose* face perceptually appears. When a prosopagnosic recognizes a face, she does so by inference from the features that *are* represented in perception. I contend that she will need to be introspectively acquainted with such perceptually represented features in order for her inference to be guided by them.

Cross-modal recognitional sorting with respect to content *is* an epistemically substantive capacity, I will now argue. I do so by looking at two alternative conceptions of our semantic self-knowledge. According to the first, the *redeployment strategy*, semantic self-knowledge is a trivial form of propositional knowledge. According to the second, *the reidentification strategy*, semantic self-knowledge is a brute form of ability-knowledge. Neither can account for our ability to cross-modally recognitionally sort.

4. The Redeployment Strategy

The literature in the philosophy of language has included extensive debate as to whether semantic externalism is consistent with self-knowledge (e.g., knowledge of the meanings

of our words).[28] According to one influential defense, suggested by John Heil (1988) and Tyler Burge (1988) and developed more fully by Christopher Peacocke (1996 & 1999 ch. 5), knowing the meanings of my words is a matter of believing certain things about them and about myself. These beliefs will "re-deploy" the very same terms whose meanings are in question. Here is a simple example: suppose that environmental factors (at least partly) fix the content of my word 'water'. Do I know what 'water' means? Sure, I do: I know that 'water' refers to water—and the very same environmental factors that fix the meaning of 'water' in my everyday usage of it fix its meaning when I reflect on the word itself. I can go on to say more about what I mean by 'water' and what I believe about water, all the while depending on environmental factors outside of my ken as the partial determinants of what my words refer to.

Heil, Burge and Peacocke have understood the redeployment strategy to apply to mental contents as well as language. So let us see if the strategy can serve present purposes, viz. whether one's *beliefs about* one's intentional mental states can make sense of one's ability to recognitionally sort them. When the potion-master's apprentice recognizes his visual experience as having the same content as his previous auditory-linguistic experience, what does the redeployment strategy have to say by way of explanation? Something like the following: the apprentice (a) forms the belief that he has heard the potion-master say thus-and-so and the belief that he is visually experiencing in such-and-such a way; and then he (b) recognizes that the beliefs have something in

---

[28] Two of the most influential papers on the side of the *inconsistency* of self-knowledge and semantic externalism come from Paul Boghossian (1989) and William McKinsey (1991). On the other end of the spectrum, Gary Ebbs has argued that if semantic externalism is true, skeptical worries about self-knowledge cannot be so much as intelligibly articulated. See Brueckner and Ebbs (2011).

common—i.e. that thus-and-so = such-and-such. But it's murky how either one of these steps is supposed to occur.

What is it for the apprentice to (a) form the requisite beliefs? It's tricky to see what *redeployment* could come to in the perceptual examples we're looking at. Notice a contrast with language: redeploying a word is a matter of using a new token of the same word-type. But what is it to redeploy a new token of the same *experience*-type? I suspect that an advocate of the redeployment strategy has two options. When the apprentice believes that he is visually experiencing in such-and-such a way, 'such-and-such' will amount to a demonstrative, filled in either by his actual visual experience (the selfsame token), or else an *imaginative recreation* of that visual experience (a new token of—sort of—the same type). (This may sound a little odd. It's natural to think that if I form beliefs about the contents of my experiences, those beliefs will have exclusively conceptual components. I will, in other words, have recognized and conceptualized the intentional contents of my experiences. But so saying would give up the dialectical significance of the redeployment strategy. The redeployment strategy is a way of explaining what such recognition comes to; it can't invoke such recognition.)

Consider an analogy with monetary currency. Coins and bills (and seashells and saltcakes and whatever else) have both intrinsic and relational properties, and their monetary value is of course among the latter. Suppose you put a coin in front of me and ask me to how much it is worth. (Let's suppose there isn't any writing on the coin that I can understand.) I tell you that it is worth *this much*, while nudging the coin forward. I think it should be clear how insubstantial my belief is. And that is as robust a belief as the redeployment strategist can reach for in accounting for the apprentice's abilities.

But now, if it is beliefs of this flavor that the apprentice forms about his two experiences, how is he supposed to (b) recognize that the beliefs have something in common? If you set a second coin in front of me, I can form a similar belief about its value. But I am not in a position to tell you, on the basis of the intrinsic[29] features of the two coins, whether the two coins have the same denomination.

The trouble here is that the apprentice's two mental states—his auditory-linguistic experience and his visual one—appear to have nothing in common *save* their intentional properties. If intentional properties are relational, then there is nothing on the basis of which sameness of intentional properties could be recognized, from the first-person perspective.

It seems, then, that the redeployment strategist is committed to there being some intrinsic property in common between mental states that have the same intentional content. But I do not see how this move is available to redeployment strategist. Here is the trouble: *there is no guarantee that content-bearing structures and intentional contents are one-to-one.* If $\$_1$ is part of a mental state, and $\$_1$ has the right functional/dispositional profile, then that state has the content F. But if $\$_2$, a different structure, has the same functional/dispositional profile, then the presence of $\$_2$ in a state makes it the case that that state also has content F. Just as two words can mean the same, or two coins can denominate the same, so content-bearing structures can instantiate the same content-fixing functional/dispositional profile. And that means that even if $\$$ *could be* part of all of one's mental states (in any mode) that have the content F (which, as I say, is already an obscure suggestion: it does not *seem* that there are intrinsic similarities between the apprentices' two experiential states distinct from their shared intentional properties),

---

[29] Not so much intrinsic as readily perceptually available, but close enough.

there is no guarantee that $ *will be* part of all of one's mental states that have the content F.[30]

Perhaps the redeployment strategist will say something like the following: there is an as-of-yet undiscovered a priori reason that content-bearing structures and intentional contents *must be* one-to-one, and hence that the apprentice is rationally justified in treating sameness of structure as an indication of sameness of intentional type. But there could not be such a reason, because it is easy to imagine systems in which structure and intentional content are not one-to-one. For example, we could set up a system such that, whenever a content-fixing state involving $_1$ is tokened, a second content-fixing state involving $_2$ is tokened in parallel. Hence $_1$ and $_2$ have the same the content-fixing functional/dispositional profile, and so any mental state that has *either one* as a part thereby has the very same content.

It seems that the best a redeployment strategist can say is that there is an as-of-yet undiscovered *empirical* reason that structures and intentional contents *happen to be* one-to-one, and hence that the apprentice is reliabilistically justified in treating sameness of activation-pattern as an indication of sameness of intentional type. (Maybe cognitive systems such as ourselves are built so that only one content-fixing state can be tokened at a time.) What I want to point out is that we have pretty much abandoned the redeployment strategy at this point. Our question was: what accounts for the apprentice's

---

[30] Further, nothing so far guarantees that a type of content-bearing structure could only admit of *one* functional/dispositional profile. Type-identical intrinsic properties could herald the presence of distinct intentional properties in distinct mental states. Just as one word can mean distinct things in distinct contexts, or one coin could denominate differently in differently social groups, so two tokens of the same structure-type could instantiate distinct content-fixing functional/dispositional profiles. It seems to me, however, that it is a constraint on naturalistic theories of intentionality that they rule out this possibility. If the same structure has multiple content-determining causal profiles, then Specificity has not been satisfied. (Not that I see *how* to rule out the possibility.)

ability to recognitionally sort his mental states according to intentional type? The redeployment strategist invoked the apprentice's same-contented beliefs about his two mental states. But we observed that the beliefs in question actually just demonstrate the mental states themselves; so the question remains how the apprentice is able to tell that the demonstrated states have the same content. The best answer so far is that the two mental states share a content-bearing structure, and that *it so happens that* sameness of structure pattern is a reliable indicator of sameness of intentional type. But the apprentice's *beliefs* about the presence of these structures are irrelevant to his recognitional ability. Furthermore, his transition from recognition of sameness of structure to sameness of intentional content is not a *rational* transition (since the one-to-one pairing of activation-pattern with intentional content is at best a contingent matter of fact). Far from semantic self-knowledge amounting to trivial propositional knowledge, such knowledge turns out (at least when it comes to knowledge about sameness and difference across mental states) to require primitive recognitional capacities whose reliability cannot be established a priori. In short: the sense in which the apprentice is "redeploying" anything is immaterial to his ability to recognize sameness and difference of intentional type, and that ability has to be epistemically brute, built into the wiring of his cognitive apparatus. The redeployment strategy has collapsed into an appeal to epistemically brute recognitional abilities. So let us look head-on at a strategy built around such an appeal.

5. The Reidentification Strategy

According to the sub-argument for premise 1, the epistemic relation that a subject bears to her intentional mental states can explain her ability to recognitionally sort them according to intentional type. It is her knowledge of them—an instance of knowing-*what*—that provides epistemic grounding for her cognitive abilities with respect to them. The advocate of the redeployment strategy rejects this: she says instead that it is a subject's beliefs about her intentional mental states (i.e. an instance of knowing-*that*) that trivially account for her cognitive abilities with respect to them. But, as we have seen, the redeployment strategy leans on primitive epistemic abilities (and doesn't do much leaning on the propositional knowledge it invokes). A second alternative reaches for primitive epistemic abilities quite self-consciously. According to this view, neither knowings-what nor knowings-that explain introspective recognitonal abilities. Rather, semantic self-knowledge *just is* a species of ability, a knowing-*how*. Here is Millikan's statement of a view of the sort I'm describing:

> The closest thing to the yearned-for ideal that actually makes some sense, I suggest, is a confrontation of one thought of an object with another thought of that same object, taking place within thought itself, and effecting a recognition of the sameness of the object.... Knowing what one is thinking of is, just, having the capacity to recognize when two of one's thought tokens are thoughts of the same.[31]

This view, which I'll call "the reidentification strategy," strikes me as phenomenologically inadequate. Here's why: my conscious attention to my intentional mental states seems epistemically relevant to my ability to recognitionally sort them. Similarly for the apprentice: if he is worried about making a mistake, he is not at a loss;

---

[31] Millikan (1993), p. 96-97.

he need not just shrug his shoulders and say, "my recognitional abilities are brute; I either have them or I do not"; he can attend extra carefully to his memory of what his Master said and to his visual experience.

The point is closely related to what John Campbell has observed with respect to the cognitive role of phenomenology in his discussion of perceptual demonstratives:

> Suppose you say to me, 'What is that mountain over there?' To understand your question I have to know which mountain you are taking about...You might acknowledge that ordinarily we would use visual information to interpret the [visual] demonstrative, but question whether it has to be conscious. The idea of visual information that is not conscious is made vivid by cases of blindsight.... The issue is whether the blindseer has the very same way of interpreting the demonstrative as the ordinary subject has. That is, the question is whether for the ordinary subject, consciousness of the object is not completely idle in an understanding of the demonstrative.[32]

The point is that if a blindseer could have the same purchase on the referent of a visual demonstrative as a normally sighted person, then visual phenomenology is explanatorily idle. But it is not: as Campbell goes on to say, "knowledge of the referent of a demonstrative comes via conscious attention to an object." My perceptual phenomenology *does explain* my ability to think about what I have seen; the comparison with blindsight serves to bring out the fact.[33]

Suppose that the potion-master's apprentice were blindsighted but nevertheless reliable in his retrieval-duties. He hears and understands his master, enters the pantry with his eyes open, and successfully reaches for the right object. Is his reliability to be explained in exactly the same way as the reliability of a normally-sighted apprentice? Clearly not: the normally-sighted apprentice is guided by his attentiveness to when his

---

[32] Campbell (2002), pp. 7-8.
[33] See Smithies (2011) for an attempt to account for the modal force of Campbell's observation. Smithies argues that not only does conscious attention *actually* play the role of delivering knowledge of the referent of a demonstrative; it is the *only possible* medium for such knowledge. The weaker, actualist observation made by Campbell is sufficient for my purposes, however.

visual-phenomenal state has a content that satisfies the description given by his master. But if the reidentification strategy were on target, the apprentice's phenomenality would make no such explanatory contribution.

The reidentification strategy does not accord with our experience of forming comparative introspective judgments, then. But phenomenological inadequacy is not the only source of trouble for the reidentification strategy; there are third-personal, theoretical reasons to reject it as well.

First, I am not certain that a system's ability to recognitionally sort mental states with respect their relational properties *could* be brute. Analogy: suppose you were charged with the task of building a robot that could sort currency of arbitrary origin for sameness of denomination. Without knowledge of the other relata in the relational property *having monetary value such-and-such*—viz, whatever states of affairs confer value on currency—there is no non-magical way to build the robot. The reidentification strategist will be forced to say—as the redeployment strategist was—that intentional mental states have intrinsic properties (such as having $ as a part) that covary one-to-one with their intentional contents. But it is at best a contingent, speculative suggestion to say that content-bearing structures and intentional contents covary in that way.

Second, while I have been arguing that the epistemic relation that a subject bears to her intentional mental states can serve to explain her recognitional abilities with respect to them, these are by no means the *only* abilities that are naturally explained by knowledge of the content of one's mental states. Here are some others such that, according to a plausible order of explanation, my knowing the content of my mental states epistemically explains them:

- my ability to knowledgeably judge that one has recognitionally sorted one's intentional mental states

- my ability to bring to mind mental states with the same content as previous mental states

- my ability to express the content of a mental state in words (e.g. describing what I want)

- my ability to express the content of a mental state in images (e.g. drawing what I see)

I now propose a dilemma for the reidentification strategist. On the one hand, she can alter her initial identification. Rather than saying that semantic self-knowledge is identical to having the single ability to recognize states with the same content, she can say that semantic self-knowledge is identical to having a *conjunction* of abilities, including those just listed. But this is an awkward adjustment. The list is disunified; nothing accounts for our including these abilities rather than others—besides the fact that we would intuitively take these abilities to all admit of the same explanation (viz. that they are rooted in one's semantic self-knowledge). And motley lists make for highly inelegant explanations. Suppose we observe that several symptoms tend to cluster together—nausea, dizziness, body aches, etc. Calling the symptom-cluster a "syndrome" is a way of naming a phenomenon that has yet to be explained; inquiry ceases only once we have discovered the underlying disease that explains the presence of all of them together. The list of capacities the reidentification strategist would have to appeal to is like this—a syndrome in need of explanation. The list is precisely the explanandum, not the explanans.

On the other hand, the reidentification strategist can retain her initial identification (of semantic self-knowledge with the ability to recognitionally sort mental states with respect to content), and then claim that *this* ability explains all the others on the list. This is probably a better route to take, but the challenge is still substantial. It is not apparent that the abilities on the list can be explained in terms of a single recognitional ability. Then again, it is not apparent that they cannot.

But there is more. According to a venerable rationalist doctrine, consciously grasping certain propositions is sufficient for knowing them, or at any rate for justifiedly believing them. Now, plenty of applications of this rationalist principle have turned out to be false. Nevertheless, it's plausible that in some cases, an adequate grasp of one's mental states does put one in a position to have justified beliefs of a certain sort—a sort we might call *conceptual* knowledge. For example, a subject's knowledge of the content of relevant mental states is sufficient to justify the following:

- the belief that twice two is four *(justified by knowledge of the content of numerically-contented mental states)*
- the belief that round things are shaped *(justified by knowledge of the content of geometrically-contented mental states)*
- the belief that poems are more like prayers than like essays *(justified by knowledge of the content of literarily-contented mental states)*

Note that if a subject were disposed to deny any of these propositions, we would consider her not just wrong but confused: we would suspect an inadequate grasp on her part of the concepts contained in them.

41

If conceptual knowledge as I have construed it is possible, it renders both horns of the dilemma more troublesome for the reidentification strategist. On the first horn, she adds such beliefs to a list of items (a subject's possession of which counts as semantic self-knowledge). But now the motley nature of the list has gotten truly out of hand: it consists of an assortment of mutually irreducible abilities *and* an assortment of mutually irreducible beliefs. The list contains elements in two different epistemic categories. Not only is there no apparent reason that the particular items are to be found on the list rather than others, there is no apparent reason why the items on the list should fall into two categories, rather than just one or several.

On the second horn, the reidentification strategist identifies semantic self-knowledge with a single recognitional ability, and then treats the abilities *and beliefs* on the list as epistemically grounded in that single recognitional ability. But it's very difficult to see how conceptual knowledge could be epistemically grounded in recognitional abilities. Supposing I am able to tell when identity-relations hold between the content of my mental states, this does not mean that I am able to tell when *other* relations hold between the contents of my mental states, relations such as determinate-determinable (as holds between roundness and shapehood).

Now, conceptual knowledge is controversial, and I haven not supplied much by way of argument in defense of my assertions about it. Specifically, I have claimed (a) that we have conceptual knowledge and (b) that our semantic self-knowledge epistemically explains it. In order to *demonstrate* the theoretical inadequacy of the reidentification strategy, I would need to defend these claims, and so doing would extend beyond the scope of the present inquiry. Consequently, the advocate of the reidentification strategy is

dialectically free to deny that we have any conceptual knowledge (maybe all beliefs are empirically justified if justified at all). But even if she were to dodge the argument from conceptual knowledge, her view would remain theoretical vicious—it would remain a list in lieu of an explanation—in addition to remaining at odds with our experience as introspective knowers.

6. Conclusion

In his 1999 book *Being Known*, Peacocke writes the following:

> Many of the philosophical problems involving the first person and the self are, in one guise or another, instances of the challenge of ... reconciling two apparent truths. The first apparent truth, an epistemological truth, is that thinkers are able to know the intentional contents of many of their own attitudes without first checking on their environmental relations. The other apparent truth, a truth of the metaphysics of mental states with intentional content, is that for a wide range of intentional contents, a thinker is, as a constitutive matter, able to have attitudes with those contents only if he stands in certain relations to an environment of a certain kind.
> In this area, the option of revising the metaphysics is unattractive. ...Anyone moved by these considerations will look for a reconciliation of our two apparent truths by reconceiving the epistemology of self-knowledge.[34]

I agree with Peacocke's description of the philosophical situation but disagree with his reconciliation-strategy. The problem is not that I *prefer* an acquaintance-based theory of self-knowledge to a belief-based or ability-based theory. The problem is that those theories are inadequate. They must either deny certain sectors of our knowledge (i.e. those that are plausibly explained by our knowledge of the content of our mental states), or treat as brute that which needs explaining. Of course an advocate of the redeployment strategy or of the reidentification strategy is free to embrace these consequences. But

---

[34] Peacocke (1999), p. 203.

there is no need to: a viable metaphysics, epistemology and semantics of *intrinsic intentionality* is available. Or so I shall argue in the chapters that follow.[35]

I conclude that the Acquaintance Argument is sound. Paradigmatic intentional properties—the sorts of intentional properties that are purely qualitative, potentially conscious and subjectively presentable—are intrinsic features of the mental states in which they are instantiated. It follows that naturalistic theories of intentional properties cannot provide a complete metaphysics of intentionality. This is a striking conclusion: it means that intentionality cannot be fully located within the ontology of the natural sciences, as that ontology is currently understood.

A final observation: I do not intend simply to dismiss the program of naturalized intentionality. True, the naturalizing project fails according to its stated aims. But it might not fail according to more modest aims: perhaps one or another of the four naturalistic theories has successfully landed on *sufficient conditions* for a mental state's having a certain content. (The mistake lies in proceeding to *identify* these conditions with having a certain content.) Or, there might be scientifically useful generalizations to be made about the properties that the four naturalistics theories identify. These properties are not the ones out of which *our* conscious, cognitive lives are built. Nevertheless it might be accurate to characterize such properties as intentional after a fashion—as the building blocks for a scientifically interesting phenomenon that we might call "proto-cognition."

---

[35] Note that I agree with Peacocke that intentional content is often environment-dependent. To the extent that this is so, empirical investigation *is* required for full semantic-self-knowledge. But it is not always so.

2. Motivating The Phenomenal Grounding Thesis

1. Introduction

If the arguments of the previous chapter have been to the point, intentional properties—at least of a certain familiar sort—are intrinsic properties of the mental states in which they are instantiated. The most promising strategies for locating intentional properties within the ontology of the natural sciences—viz., so-called "naturalized" theories of intentionality—try to reduce intentional properties to functional/dispositional profiles. But conscious mental states cannot by themselves underwrite these functional/dispositional profiles: only mental states *in tandem with their embedding cognitive systems and/or environments* can do so. N-I properties—the properties picked out by these functional/dispositional profiles—are therefore *non-intrinsic* to the mental states that include them. Hence, there is more to intentionality than N-I properties.

But if intentional properties are not to be accounted for in terms of the functional/dispositional profiles of the mental states in which they are instantiated, what resources might we use to account for them? A promising (if underdeveloped) answer can be found in the work of a handful of dissenters from the naturalized intentionality paradigm. These philosophers have suggested that there is a close relationship between content and consciousness, or between intentional properties and phenomenal properties. Some have gestured at an asymmetric explanatory relationship between them: for example, that a state has the intentional properties it has "in virtue of" its phenomenal character, that its phenomenal character "constitutively determines" its intentional

content, or that phenomenality is the "basis" of mental content.[1] In this chapter and the

next, I explore a thesis that makes the asymmetric explanatory relationship explicit and

precise, a thesis I will call "The Phenomenal Grounding Thesis":

> **PGT**: All instances of intentional properties are partly grounded in instances of
>
> phenomenal properties.

The literature on the relationship between consciousness and content contains a number

of arguments that have motivated views in the neighborhood of PGT. In this chapter I

reconstruct and evaluate the most influential of these arguments. In the next chapter I

sketch my preferred version of the metaphysics underlying PGT and respond to some

objections to PGT.

2. Perceptual Phenomenal Sufficiency

PGT expresses an explanatory asymmetry. Consequently, motivating it requires

establishing two claims: first, that there is a tight connection between explanans and

explanadum; second, that the connection is asymmetric. I will briefly discuss the second

claim in the next chapter. I will devote the remainder of this chapter to the first claim.

Specifically, I will be discussing what we might call the "Inseparability Thesis": For a

class of intentional properties and a class of phenomenal properties, a subject instantiates

items in the first class iff she instantiates items in the second class. (I will be assuming

that the modal strength of the Inseparability Thesis is metaphysical necessity, though the

relevant literature does not always make the matter clear.) As stated, the Inseparability

---

[1] The first is Uriah Kriegel (2011), the second Terry Horgan and John Tiensen (2002), the third Brian
(2003b).

Thesis admits of considerable variation depending on how the classes get specified and on other interpretive factors.

I'll begin by discussing the left-to-right conditional. Here is another way of putting the claim: there are phenomenal properties such that, necessarily, if a subject instantiates them, she thereby instantiates certain intentional properties as well. Consciousness is, in certain forms, *sufficient for* intentionality. Call this idea "Phenomenal Sufficiency". There is a less controversial version of phenomenal sufficiency and a more controversial version (though I will be arguing that the more controversial version really should not be). According to the less controversial version, some perceptual states include phenomenal properties, such that, necessarily, if a subject instantiates them, she instantiates certain intentional properties as well. Call this thesis "Perceptual Phenomenal Sufficiency", or PPS. According to a more controversial version, some non-perceptual states include phenomenal properties, such that, necessarily, if a subject instantiates them, she instantiates certain intentional properties as well. Call this thesis "Non-Perceptual Phenomenal Sufficiency," or NPS.

Before discussing the arguments for PPS, it will be helpful to briefly sketch the most common views about the relationship between intentionality and phenomenology in perception. We can divide these views into those that are consistent with PPS, which I'll call "inseparabilist" views, and those that are inconsistent with PPS, which I'll call "separabilist" views. Inseparabilist views divide into *phenomenology-first inseparabilism*, and *intentionality-first inseparabilism*. According to phenomenology-first inseparabilism, some intentional properties of perceptual states are grounded in those states' phenomenal properties. (This is the view I will be developing in subsequent

47

chapters.) According to intentionality-first inseparabilism, some phenomenal properties of perceptual states are grounded in those states' intentional properties, and in no other way. (The qualification 'and in no other way' is needed to rule the possibility that phenomenal properties can be grounded in more than one way, one of which does not involve intentional properties. This would be inconsistent with PPS.)

There are two common forms of intentionality-first inseparabilism. The first is *reductive representationalism*, according to which perceptual phenomenology is reducible to representational content, or to representational content in connection with a perceptual state's functional role. (If there can be non-conscious representational states, or representationally equivalent states in more than one perceptual mode, then representational states alone cannot ground perceptual phenomenology.) The second is *direct realism*, according to which perceptual states are relations between a subject and that which the subject perceives. On direct realism, a subject's relation to what she perceives constitutes both perceptual phenomenology and perceptual intentionality. I mention direct realism only to set it aside: it honors the letter of PPS but not its spirit. Direct realists maintain that perceptual intentionality supervenes in part on features of the subjects' environment, and this is a core tenet shared by PPS's detractors. To say that PPS is true because perceptual phenomenology likewise supervenes on features of the subjects' environment is, I think, to gut PPS of most of its philosophical interest. Inasmuch as the Argument from Discriminability discussed below puts pressure on perceptual externalism (the idea that all perceptual intentionality partly supervenes on features of the subjects' environment), it likewise puts pressure on direct realism.[2]

---

[2] The very same considerations that I will use to put pressure on perceptual externalism count against the direct realist conception of phenomenology. By identifying phenomenology with subject-world relations,

48

There are two important separabilist views. The first is *pure qualia theory*. According to pure qualia theory, perceptual states include intentional properties only contingently. The essential nature of perceptual states consists in "pure" (i.e. non-intentional) perceptual sensations, which give rise to or are accompanied by perceptual *judgments* (which are, of course, essentially intentional). The second is *perceptual externalism*. According to perceptual externalism, perceptual states include intentional properties essentially, but which intentional properties are included does not supervene on which phenomenal properties are included. Pure qualia theorists and perceptual externalists do not agree on the nature of perceptual states, but they do agree that perceptual intentionality is never intrinsic to perceptual phenomenology—and for purposes of understanding the dialectic surrounding PPS, such agreement warrants lumping them together. I'll call such opponents of PPS "Externalists."

Arguments in the literature for PPS can be consolidated into three types: the Argument from Phenomenal Duplication, The Argument from Intentional Contrast, and the Argument from Discriminability.

The Argument from Phenomenal Duplication. The following thought experiment comes from Brian Loar (2003): suppose that, while participating in a psychology study, you undergo a series of visual experiences as of indistinguishable lemons. You are told afterward that some of your experiences were induced hallucinations. Though the *causes* of your type-identical experiences differ, their *accuracy-conditions* (and hence intentional

direct realism entails that the elements of a phenomenal state can be related to the subject no more closely than via a causal relation. This would rule out the possibility of the subject's being introspectively acquainted with elements of her own phenomenal states. As I argued in section 3 of the last chapter, judgments made on the basis of introspective acquaintance are not subject to appearance/reality fallibility, but judgments made about items related to the subject no-more-closely-than-causally *are* subject to appearance/reality fallibility. If we add here that no element of a phenomenal state is in principle beyond the purview of introspective acquaintance, it follows that direct realism is wrong about the nature of perceptual phenomenology.

contents) do not. More dramatically[3]: a life-long envatted duplicate of your brain instantiates all of your phenomenal states. It is plausible that the envatted subject's perceptual states are systematically non-veridical, and the subject's phenomenology—which by hypothesis is identical with yours—is sufficient to establish such systematic non-veridicality. According to the Argument from Phenomenal Duplication, the reason it seems to us that phenomenal duplicates must share intentional contents is that some intentional content supervenes on the phenomenal character of conscious perceptual states. Put more formally:

(1) We cannot conceive of pairs of perceptual states, P1 and P2, such that (a) they do not differ in any phenomenal properties, but (b) they differ in all intentional properties.

(2) If (1), it is impossible for P1 and P2 (a) not to differ in any phenomenal properties but (b) to differ in all intentional properties.

(3) Hence, some perceptual states include phenomenal properties, such that, necessarily, if a subject instantiates them, she instantiates certain intentional properties as well.

The Argument from Intentional Contrast. Whenever we perceptually resolve a visually ambiguous image—e.g. see the image as a duck vs. a rabbit, as a young woman vs. an old woman, a concave vs. a convex cube etc.—our visual experience undergoes simultaneously phenomenal and intentional changes. Similarly, the *intentional* difference between a visual agnosic (who cannot visually recognize familiar objects) and someone who has normal visual recognitional capacities is accompanied by a *phenomenal*

---

[3] Loar (2003b), Horgan & Tienson (2002), and Graham, Horgan & Tienson (2004) all lean on BIV thought-experiments.

difference.[4] According to the Argument from Intentional Contrast, the reason it seems to us that such intentional contrasts as these case depict must be accompanied by phenomenal contrasts is that some intentional content supervenes on the phenomenal character of conscious perceptual states. Put more formally:

(1) We cannot conceive of pairs of perceptual states, P1 and P2, such that (a) they differ in the intentional properties presented to the subject therein, but (b) they do not differ in any phenomenal properties.

(2) If (1), then it is not possible for P1 and P2 (a) to differ in the intentional contents presented to the subject therein, but (b) fail to differ with respect to their phenomenal character.

(3) Hence, some perceptual states include phenomenal properties, such that, necessarily, if a subject instantiates them, she instantiates certain intentional properties as well.

So formulated, the Arguments from Phenomenal Duplication and Intentional Contrast remain schematic: filling them in would require specifying the types of states that count as values for P1 and P2. Such specifications could differ with respect to *scope* (i.e., which sorts of states exemplify the supervenience of intentional content on phenomenal character) and *grain* (at what level of determinacy exemplifies the supervenience of intentional content on phenomenal character). But I take that PPS is interesting even if it has very few witnesses. The suggestion that intentional content *ever* supervenes on phenomenology (let alone *often* does so) is by itself very interesting metaphysically, and likely to be challenged by Externalists. So we can leave the scope and grain of P1 and P2 unspecified for present purposes.

---

[4] These examples come from Bayne (2009). Siegel (2010) also makes central use of contrast arguments.

I doubt that Externalists are going to be moved by these arguments. Likely they will deny (1) in each argument—viz., the premise that declares inconceivable the severing of certain phenomenal/intentional pairings. (If they don't deny (1), they'll certainly deny the inference from (1) to (2)—viz., the inference from inconceivability to impossibility.) On what sort of picture would it make sense to do so? I think the most natural thing for the Externalist to say in response to the Argument from Phenomenal Duplication is that perceptual content is essentially *object-involving.* Thus, it's a mistake to think that phenomenal duplicates share any perceptual contents—even if what it's like from the subject's point of view is the same in each case. Now, it's worth asking whether this response makes sense. Do duplicate states P1 and P2 really share *no* intentional contents? Might they, for example, share *existential contents* ('there exists a lemon yonder') or *modes of presentation* of object-involving contents ('that yellow ovoid thing')? The Externalist will have to answer no to these questions. I leave it to reader to judge whether such an answer is plausible.[5]

The Externalist will need a different strategy for denying (1) in the Argument from Intentional Contrast, because (by hypothesis) the relevant intentional contrasts do not involve any changes in which objects are perceived. A natural strategy would be to grant that the relevant intentional contrasts do typically involve changes in phenomenology, but to maintain that these changes are only contingently correlated with intentional changes. Carutthers & Veillet (2011) take up this strategy: "Subjects who wear eye-trackers while viewing ambiguous figures show different patterns of overt

---

[5] Lycan (2008, p. 243) writes: "In each case, a particular lemon is represented. Some of the lemons are real, and some are unreal, but that does not change the fact that in each case you visually demonstrate an identical lemon. The experiences are indistinguishable because the lemons themselves are indistinguishable." I find this appeal to unreal particulars mystifying.

attention under the two viewing conditions...It is hardly surprising, then, that the

phenomenology of one's experience should differ in the two cases. This is because the

concepts that one deploys have a causal impact, via patterns of attention."[6] Gestalt-shifts,

that is, are accompanied by changes in perceptual attention, and changes in perceptual

attention are phenomenally felt.[7] Similarly, Carutthers & Veillet grant that there is a

phenomenal difference between agnosic perception and normal perception, but maintain

that the difference can be explained in terms of the feelings of effort and frustration that

the agnosic undergoes while struggling to categorize. Now, since the connection between

intentional contrast and phenomenal change is a contingent one, Carutthers & Veillet will

presumably deny that it is *inconceivable* that the connection could break down—and

hence that (1) in the Argument from Intentional Contrast is false.

    The Argument from Discriminability. There are ways for Externalists to resist

Arguments from Phenomenal Duplication and Intentional Contrast, then. This doesn't

mean that those arguments are dialectically worthless; they may very well persuade non-

partisans who find the Arguments' premises more intuitive than they find Externalism's

rebuttals. But a third type of argument does have the potential to put pressure on

Externalism. According to this argument, our discriminatory knowledge with respect to

the contents of perception can only be explained if differences in contents show up as

differences in phenomenal character. We find a sketch of such an argument in Siegel

(2010). Her target is pure qualia theory, according to which "neither objects nor

properties are presented in experience, which leaves it mysterious what role experience

---

[6] (Carruthers & Veillet 2011, p. 40).
[7] They write, "Subjects who wear eye-trackers while viewing ambiguous figures show different patterns of overt attention under the two viewing conditions...It is hardly surprising, then, that the phenomenology of one's experience should differ in the two cases. This is because the concepts that one deploys have a causal impact, via patterns of attention" (Carruthers & Veillet 2011, p. 40).

plays in enabling the subject to distinguish objects from one another and figure from ground."[8] I think her argument can gain traction not just against pure qualia theory but against perceptual externalism, too—since both types of Externalism deny that intentional content is intrinsic to perceptual phenomenology. Here is my gloss on Siegel's argument:

(1) It is possible for a subject S to discriminate objects on the basis of the features presented in her conscious perceptual states.

(2) If (1), then S can be introspectively acquainted with the intentional contents presented in her conscious perceptual states.

(3) If S can be introspectively acquainted with a feature X presented in her conscious perceptual states, then there is phenomenal feature Y such that X supervenes on Y.

(4) Hence, some perceptual states include phenomenal properties, such that, necessarily, if a subject instantiates them, she instantiates certain intentional properties as well.

The thought here is that a subject's introspective acquaintance with the intentional contents of her conscious perceptual states can epistemically explain her discriminatory abilities with respect to them. Hence intentional contents must somehow be trackable by the subject via introspection. If intentional contents failed to supervene on the phenomenal character of her perceptual states, then this would not be possible.

This is familiar territory: the argument from discriminability is closely related to the Acquaintance Argument from the previous chapter. The present argument goes one step further than the Acquaintance Argument in identifying the *medium* of introspective

---

[8] Siegel (2010), p. 72.

acquaintance with phenomenality. Plausibly, for a subject to be introspectively

acquainted with some item, that item has to *appear* somehow to the subject—and for

something to appear somehow to the subject is just for it to have phenomenal character.

Premise (1) of the argument from discriminability is just as contentious as is

premise (1) of the Acquaintance Argument: it is open to the perceptual externalist to treat

S's discriminatory abilities as epistemically brute. As we saw in chapter 1, such a

position is phenomenologically inadequate, since it suggests that a subject whose

perceptual states were all unconscious could be in an epistemic situation just as good as a

conscious perceiver, with respect to discriminating the objects in her environment. If

that's what the Externalist has to say, then there is not much to be said for Externalism.

So the case for PPS is strong.

3. Non-Perceptual Phenomenal Sufficiency

Granting PPS leaves open the question of whether intentional properties supervene on

phenomenal properties in conscious states outside of perception—in other words,

whether Non-Perceptual Phenomenal Sufficiency (NPS) is true. The dialectic

surrounding NPS is very similar to the dialectic surround PPS. That is, two types of

views are consistent with it: phenomenology-first inseparabilism and intentionality-first

separabilism.[9] (The first is the view that some intentional properties of non-perceptual

states are grounded in those states' phenomenal properties. The second is the view that

some phenomenal properties of perceptual states are grounded in those states' intentional

---

[9] I'm not aware that anyone espouses a corollary to direct realism for non-perceptual states. I gather that
this is because non-perceptual states are so routinely about non-actual states of affairs. I can, for example,
run through a number of possibilities for tonight's main dish. A corollary of direct realism would explain
this fact in terms of some sort of relation I bear to merely possible states of culinary affairs.

properties, and in no other way.) Inconsistent with NPS are separabilist views akin to pure qualia theory and phenomenal externalism, but applied to non-perceptual states. (The first is the view that non-perceptual states have a phenomenology but need not have content; the second is the view that non-perceptual states need to have a phenomenology and some content or other but there is no modally strong guarantee of *which* content.) I will again lump these views together under the heading 'Externalism.'

A number of philosophers who grant PPS have been more skeptical of NPS. Why is NPS more controversial? I think there are two reasons. The first reason is phenomenological. While there is disagreement about the phenomenological data regarding perceptual states, such data is relatively easy to come by, when compared with imaginative and cognitive states. Not only is there extensive debate about what such states are like, there is even debate about whether such states are like anything at all.[10] The second reason is empirical. According to Ray Jackendoff (1987), the best account of the relationship between cognition, perception and consciousness locates consciousness at an intermediate level of perceptual processing. Low-level perceptual features (colors, shapes, and so forth) are phenomenally represented but higher cognitive processing happens "above" the level of consciousness in the hierarchy. Those "restrictivists" (the term is Jesse Prinz's) who endorse Jackendoff's picture are not likely to endorse a close relationship between consciousness and content for a very broad range of mental states. In what follows I hope to show that it is not so easy to maintain PPS while denying NPS; very similar considerations count in favor of both.

---

[10] It is an interesting question why some phenomenal domains yield a lot of consensus about their nature and structure whereas other domains yield very little. I suspect it has to do with whether states in a domain are readily *attendable*.

Arguments in the literature for NPS can be consolidated into the same three types as above: the Argument from Phenomenal Duplication, The Argument from Intentional Contrast, and the Argument from Discriminability.

The Argument from Phenomenal Duplication. Suppose that Externalism about non-perceptual states were true: the phenomenal feel of a conscious thought would have no implications for the content of that thought. Charles Siewert (1998) finds the idea inconceivable: "The suggestion would be that it might seem to one *just* the way it does to me to *say*, 'Bob is driving home from work now,' and to *think* that Bob is driving home from work now, even when one did not have this thought, *or any thought at all.*"[11] According to the Argument from Phenomenal Duplication, the reason it seems to us that phenomenal duplicates must share intentional contents is that some intentional content supervenes on the phenomenal character of conscious non-perceptual states. Put more formally:

(1) We cannot conceive of pairs of non-perceptual states, N1 and N2, such that (a) they do not differ in any phenomenal properties, but (b) they differ in all intentional properties.

(2) If (1), it is impossible for N1 and N2 (a) not to differ in any phenomenal properties but (b) to differ in all intentional properties.

(3) Hence, some non-perceptual states include phenomenal properties, such that, necessarily, if a subject instantiates them, she instantiates certain intentional properties as well.

The Argument from Intentional Contrast. Some cases of conscious gestalt-switching are such that explanation in terms of *perceptual* changes seems inadequate.

---

[11] Siewert (1998), p. 284.

Central cases are those that involve contrasts in the bits of language are understood.

Galen Strawson (1994), for example, points to the contrast between hearing a language

one does not understand vs. hearing a language with understanding.[12] Similarly, Siewert

(1998) and Horgan & Tiensen (2002) discuss cases of cognitively resolving an

ambiguous sentence: "imagine hearing or saying 'Time flies' as a cliché about the

passage of time, vs. saying or hearing it as a command at the insect races. The actual

sound or auditory imagery may be the same, but the total experiences are phenomenally

quite different."[13] Siewert describes similar cases that do not involve any sort of

perception, e.g., sudden eruptions of wordless, imageless thoughts.[14] According to the

Argument from Intentional Contrast, the reason it seems to us that such intentional

contrasts as these case depict must be accompanied by phenomenal contrasts is that some

intentional content supervenes on the phenomenal character of conscious non-perceptual

states. Put more formally:

(1) We cannot conceive of pairs of non-perceptual states, N1 and N2, such that (a)

they differ in the intentional properties presented to the subject therein, but (b)

they do not differ in any phenomenal properties.

(2) If (1), then it is not possible for N1 and N2 (a) to differ in the intentional contents

presented to the subject therein, but (b) fail to differ with respect to their

phenomenal character.

(3) Hence, some perceptual states include phenomenal properties, such that,

necessarily, if a subject instantiates them, she instantiates certain intentional

properties as well.

---

[12] Strawson (1994) p. 5ff.
[13] Horgan & Tienson (2002) p. 523. Cf. Siewert (1998) p. 279.
[14] Siewert (1998) pp. 276-77.

How might Externalists respond to these arguments? As above, I think they will deny (1)
in each argument. Now, in connection with the Argument from Phenomenal Duplication
for PPS, I said that the Externalist will likely appeal to object-involving content, which
phenomenal duplicates do not share. This is probably not a good strategy for the
Externalist to take in the present context, however, since the content of imaginative and
cognitive states can involve *generic* particulars (I can imagine *a* lemon without imagining
any lemon in particular) and *nonexistent* particulars (I can imagine a cow-shaped lemon).
A different strategy is available, though. The Externalist can maintain that there is no
specific phenomenology that answers to the description, *What it's like to think "Bob is*
driving home from work now". Rather, to say that there is something it is like to think
'Bob is driving home from work right now' is to say that such a thought is contingently
accompanied by phenomenology that fits or suits it, such as inner speech or visual
imagery as of Bob in his car. Thus, while there is a close connection between the content
and the phenomenology of the conscious thought *Bob is driving home from work now*,
that connection is not *necessary*, and we can conceive of its breaking down.

This sort of strategy can also be pursued as a way to resist the Argument from
Intentional Contrast. First of all, the Externalist can note the affinities between examples
of intentional contrast given for non-perceptual states, and examples of intentional
contrast given for perceptual states above. That is: the shift from not understanding to
understanding an utterance, or from parsing it one way to parsing it another, amount to
types of perceptual gestalt-shift. So perhaps concomitant shifts in attention can explain
the relevant phenomenal changes, just as (the Externalist maintains) they explain
phenomenal changes in the resolution of ambiguous figures and so forth. Second, where

these affinities with perception break down (as in the case of sudden wordless, imageless thoughts), the Externalist can point to changes in concomitant imagery, in keeping with her strategy for responding the Argument from Phenomenal Duplication.[15] Tye & Wright (2011) demonstrate how to put this dual strategy to use:

> When we hear someone speaking in a language we do not understand, the phonological processing of the sound stream is different from the processing that goes on when we hear someone speaking in a language we do not comprehend. In the former case, the 'grouping' of the sound stream is causally influenced by the semantic processing. The result is that the auditory experiences we undergo are different from those we would undergo were we to hear the same sound stream without understanding the language.[16]
> Hearing [the sentence 'Time flies'] one way, there is a phenomenal difference from hearing it the other way, [Horgan & Tienson] claim. We agree. But so what? Isn't the difference easily accounted for by differences in associated linguistic images or auditory experiences having the phonological and syntactic structure of the sentence 'time flies' under each interpretation (and perhaps differences in visual images too—of flies lined up to compete, say, in the latter case but not in the former)?[17]

(What to say about sudden eruptions of wordless, imageless thoughts? Though Tye & Wright don't address the matter directly, I think that the natural thing for them to say is that if there are any, they are accompanied by some *affective* shift—just as Carruthers & Veillet maintained above regarding agnosic vs. normal perception. Specifically, one suddenly has feelings of alarm, or insight, or possibility, or something similar.)

I have a quibble and a criticism for this strategy. Here is the quibble. Tye & Wright presumably mean to appropriate phonological and syntactical contrasts to *attentional* contrasts. But it is not obvious that they can be so appropriated. True, words, phrases, and sentences do seem to serve as units of auditory attention. The question is whether they first need to be recognized as words, phrases, and sentences in order for

---

[15] The general shape of this strategy goes back to Lormand (1996). It gets developed by Caruthers & Veillet (2011), Prinz (2011), Tye & Wright (2011), and William Robinson (2005 & 2011).
[16] Op. cite. p. 337.
[17] Ibid.

60

them to serve as attentional units. Compare: in visual perception, objects can serve as the units of visual attention—precisely because they appear to us *as objects*, as already differentiated from one another and from the background. If we have to represent objects as objects before we can attend to them, then it's plausible that we have to represent phonemic groupings as phonemic groupings before we can attend to them. Attentional shifts alone can't explain the phenomena: what's involved is a type of perceptual recognition, an attribution of a feature (*being a phoneme*) to an auditory percept.[18] So saying is still consistent with denying NPS: we're still just talking about a kind of perceptual representation. But it isn't consistent with denying PPS. Tye & Wright's strategy for denying NPS presupposes that PPS is true.

Here is the criticism. the appeal to imagery is legitimate only if imaginative states are not intrinsically intentional. Consider what is involved in visually imagining a fly-race: it is not as though an undifferentiated blur of colors and shapes comes before your mind. If that were so, identifying your mental image *as an image of a fly-race* would require a further act of disambiguation. But that is clearly *not* so. Try to conjure an image of the fly-races: now, is it possible that your imagery is actually of the *duck*-races, but the contestants are wearing fly-costumes? Is it possible that your imagery is actually of the filming of the fly-race *documentary*? Of course not.[19] Nor do these appeals to what is intuitively obvious need to carry the day, because we could easily construct a version of the Argument from Discriminability for imaginative states. If anything, it seems *more* clear that the intentional contents of imaginative states are intrinsic to them, than that the intentional contents of perceptual states are intrinsic to them. It turns out that the appeal

---

[18] See O'Callaghan (2010) and (2011) for interesting discussion of these matters.
[19] McGinn (2005) argues persuasively that imaginative episodes can be essentially intentional.

to imagery *commits* the Externalist to NPS, rather than supplying her with a means of denying NPS.

Now, none of this would matter if the Externalist could avoid appealing to imagery in pushing back against the Arguments from Phenomenal Duplication and Intentional Contrast. But it is very hard to see how she could do so. These Arguments purport to show that intentional content supervenes on phenomenal character in cognitive states. The appeal to imagery is a way of denying this: same-contented cognitive states tend to be accompanied by identical imagery, different-contented cognitive states tend to be accompanied by contrasting imagery, and that is all there is to it. But if the appeal to imagery is illegitimate, then the Externalist will have to rest her response on a yet more primitive form of phenomenology as the contingent accompaniment of cognitive content. What form could this be? I have no idea.

The Argument from Discriminability. NPS is very likely true, at least when it comes to *imaginative* states. A final argument purports to show that there must be "purely cognitive phenomenology": the phenomenology as of thinking particular thoughts. David Pitt has argued at length that there is such phenomenology. Here is his original formulation of his argument:

> Normally—that is, barring confusion, inattention, impaired functioning, and the like—one is able, consciously, introspectively and non-inferentially (henceforth, "Immediately") to do three distinct (but closely related) things: (a) to distinguish one's occurrent conscious thoughts from one's other occurrent conscious mental states; (b) to distinguish one's occurrent conscious thoughts each from the others; and (c) to identify each of one's occurrent conscious thoughts as the thought it is (i.e., as having the content it does). But (the argument continues), one would not be able to do these things unless each (type of) occurrent conscious thought had a phenomenology that is (1) different from that of any other type of conscious mental state (proprietary), (2) different from that of any other type of conscious

thought (distinctive), and (3) constitutive of its (representational) content (individuative).[20]

Pitt contends, that is, that *the only explanation* of our discriminatory abilities with respect to our occurrent cognitive states is that their intentional content supervenes on their phenomenal character. Why should we agree? I think the best way to reconstruct Pitt's argument appropriates it to the same form as the version previously discussed in the context of PPS, viz.:

(1) It is possible for a subject S to discriminate her occurrent thoughts on the basis of the features presented in her conscious perceptual states.

(2) If (1), then S can be introspectively acquainted with the intentional contents presented in her conscious cognitive states.

(3) If S can be introspectively acquainted with a feature X presented in her conscious cognitive states, then there is phenomenal feature Y such that X supervenes on Y.

(4) Hence, some non-perceptual states include phenomenal properties, such that, necessarily, if a subject instantiates them, she instantiates certain intentional properties as well.

The crucial premise in the argument is, of course, premise (1), and Pitt does nothing to support it; an opponent might propose instead that our discriminatory capacities are brute, though this would again violate phenomenological adequacy: it would amount to the implausible suggestion that a subject whose cognitive states were all unconscious could be in an epistemic situation just as good as a conscious thinker, with respect to discriminating the contents of her thoughts. Alternatively, one could propose that intrinsically-intentional imagery provides the phenomenological resources with which to

---

[20] Pitt (2004), p. 7.

discriminate occurrent thoughts. Whether this proposal is adequate will turn on whether (a) non-imagistic conscious thought is possible, and (b) whether non-imagistic conscious thoughts are discriminable in the same way that image-accompanied conscious thoughts are. We will have occasion to revisit these matters in chapter 5. Whether or not we go in for pure cognitive phenomenology, the existence of intrinsically-intentional imagery is enough to demonstrate the truth of NPS—and that is good enough for present purposes.

4. Phenomenal Necessity: Subjectivity.

I have been discussing the "Inseparability Thesis": For a class of intentional properties and a class of phenomenal properties, a subject instantiates items in the first class iff she instantiates items in the second class. I now turn to the right-to-left conditional: there are intentional properties such that, necessarily, if a subject instantiates them, she instantiates certain phenomenal properties as well. Consciousness is in some sense *necessary for* intentionality. Call this idea "Phenomenal Necessity".

I will begin by briefly discussing two strongly restricted versions of Phenomenal Necessity, before turning my attention to more expansive versions. (When I say "restricted" and "expansive," I now shift my attention from the class of phenomenal properties to the class of intentional properties, i.e. the extent of intentional phenomena for which consciousness is necessary.) First: we have thoughts *about* our own phenomenal states. We can, for example, reflect on the difference between the appearance of a periwinkle thing against a blue vs. a purple background; we can remember what it was like to have a tooth anesthetized by the dentist; we can form a judgment about our current mood. Now, it is controversial how to characterize the

phenomenal concepts embedded in such thoughts, but almost everybody agrees that

possession of such concepts requires having experienced the relevant phenomenal

property.[21] (The agreement seems to cut across the divide between physicalists and anti-

physicalists, and between those who hold that phenomenal concepts are bare

demonstratives, on the one hand, and those who hold that phenomenal concepts have a

richer connotation, on the other.[22]) Given that thoughts about our own conscious states

have non-trivial truth values—that is, assuming that they are genuinely descriptive as

opposed to merely expressive, and that they are not self-verifying—then it follows that, at

the very least, our having *consciousness-directed thoughts* entails that we are conscious.[23]

According to a second, somewhat less restrictive, version of Phenomenal

Necessity, consciousness is necessary for entertaining perceptual-demonstrative thoughts.

This is one of the central theses for which John Campbell argues in his book *Reference*

and Consciousness. He starts from the assumption that knowledge of the referent of a

demonstrative is what causes and justifies further investigative activity vis-a-vis

perceived objects. Further, conscious perceptual attention is what makes such knowledge

possible. He illustrates: "consider an ordinary cases in which you and I are sitting at a

dinner table with a large number of people around and you make a remark to me about

'that woman'...so long as my conscious experience remains a sea of faces, there is an

---

[21] As I understand him, Kirk Ludwig is a dissenting voice: he has observed (in conversation) that one can *acquire* a concept any which-way (by quantum accident, perhaps), so concept-possession should not ever be cashed out in terms of canonical methods for concept-*acquisition*. It might nonetheless turn out that possessing a phenomenal concept at least requires *being disposed to experience* the relevant phenomenal property.

[22] Representative advocates include Levine (2007), Papineau (2002), and Tye (1999).

[23] It probably also entails that consciousness is not an epiphenomenon, assuming that the relation between concrete particulars, on the one hand, and justified belief about them, on the other, has to be at least as close as causality.

ordinary sense in which I do not know who you mean."[24] Supposing Campbell is right

about the data, what explains the fact that conscious attention is necessary for perceptual-

demonstrative reference? His answer is that attention plays the role of selecting, from

among the plenitude of information presented in experience, those parameters that fix

demonstrative reference.

Campbell's case is in one sense stronger than he makes out and in another sense

weaker. It is stronger in the sense that, if he is right that conscious attention is necessary

for demonstrative reference, then consciousness is not only necessary for perceptual-

demonstrative thought. It's also necessary (causally, anyway) for thoughts that deploy

concepts acquired via the use of perceptual demonstratives. It is obvious that vast

quantities of our conceptual repertoire are acquired in this way. But Campbell's case is

weaker than he makes out because he fails to show that conscious attention performs the

relevant selection-task. For all he says, it could be that conscious attention

epiphenomenally *accompanies* non-conscious selection-tasks. As Declan Smithies has

convincingly argued[25], what is needed to motivate Campbell's position is a just-as- (or

still-more-) controversial claim in epistemology, viz. phenomenal internalism about

justification. That is, if (1) demonstrative judgments are among the immediately-

perceptually justified judgments; if (2) demonstrative reference requires attending to a

subset of the perceptually-available information; *and* if (3) the determiners of

justification must be accessible to consciousness, then perceptual attention must be (at

least potentially) conscious. But claim (3) is highly contentious, so Campbell's case is

hard to evaluate.

---

[24] Campbell (2002), p. 9.
[25] Smithies (2011).

I turn now to three ways of defending a much more expansive Phenomenal Necessity claim. According to these lines of thought, a great many of our intentional states depend for their existence, in some sense, on our capacity for phenomenality. According to the first, consciousness is necessary for the *subjectivity* of intentional states; according to the second, consciousness is necessary for the *determinate content* of intentional states; according to the third, consciousness is necessary for the *cognitive unity* among our intentional states. All three lines of argument can be traced to John Searle, though Searle has by no means offered the last word on any of them. I will focus on the first of these three ways of defending Phenomenal Necessity in what remains of the present section.

Searle uses his famous "Chinese Room" thought experiment to argue that the functional facts are insufficient to fix the intentional facts.[26] Searle asks us to imagine that we have been assigned the following task: when given some writing in Chinese as "input" (assuming we do not know a word of Chinese), we are to follow a list of rules written in English that correlate inputs and outputs, and finally to copy the Chinese characters mentioned in the rules "as outputs." No matter how proficient we were to become at processing these inputs and outputs and no matter how closely they mimic meaningful Chinese discourse, it is natural to say that neither we nor the functional system of which we form a part *understand Chinese*; at the very least it is an open question whether the whole operation attains to genuine understanding. Searle concludes that there is an explanatory or conceptual gap between symbol-processing that conforms to rules, on the one hand, and semantic understanding, on the other.

---

[26] Searle (1980).

What is missing? Searle does not say (in the same paper, at any rate). One possibility can be found in the writing of Colin McGinn, who has also pointed to what he sees as an explanatory gap between functional states of an organism and its intentional mental states. McGinn thinks that, even once we include *features of the system's environment* and the *homeostatic needs of the system* in the picture (in addition to the computational states that Searle focuses on), we are still lack an explanation for how the system could genuinely understand. He writes, "We can ask ourselves whether it really seems plausible that any of the standard theories capture the complete nature of conscious intentionality. ... They do not capture that phenomenological feature we describe (somewhat metaphorically) as grasping, apprehending, reaching out, taking in, and so forth."[27] For a system to be in an intentional state, McGinn goes on to suggest, is not just a matter of that states' having an intentional object but also of its having an intentional *subject*. My conscious intentional states are *mine*, they are *for me*. A conscious intentional state is not just directed at the world but is also "present to the subject;" it is "Janus-faced."

I suspect that Searle has something similar in mind. One of his repeated criticisms of theories of intentionality that make no mention of consciousness is that they cannot capture a phenomenon he calls "aspectuality": "Intentional mental states represent their conditions of satisfaction only under certain aspects and those aspects *must matter to the agent*" (emphasis mine).[28] (Searle's notion of aspectuality is thus a conjunctive notion; the second conjunct is relevant to the present discussion, whereas the first conjunct will be relevant the argument from content determinacy, which we will examine next.) The

---

[27] McGinn 1988, pp. 240-41.
[28] Searle 1991, p. 50.

Searle-McGinn view is something like this, then: some intentional states have the feature of *being presented to a subject*, and the functional facts are insufficient to account for this feature.[29] In contrast to the "outward-directedness" of intentional properties, I'll call this feature "inward-directedness," or "inwardness," and I'll call the Searle-McGinn view "the inwardness intuition."

Here are two clarificatory questions about the inwardness intuition: first, what is the connection supposed to be between an intentional state and the property of inward-directedness? Second, how might the inwardness intuition support Phenomenal Necessity? An answer to the second question will help with the first, so I will start there. It is natural to think that there is a tight connection between consciousness and subjectivity, though it is hard to pin down exactly what the connection is. Consciousness, we might say, is the medium of subjectivity; there is nothing of which a subject is aware that is not presented in or by consciousness. Perhaps this is because subjectivity *just is* consciousness—though it is hard to know what such a claim comes to in the absence of a theory of one or the other. Perhaps, instead, this is because phenomenal properties can only be instantiated by a subject (just as a texture can only be instantiated by a surface, say). However the ontology of the mysterious—and consequently often finessed—topic of subjectivity gets worked out, it is at least plausible to say that subjectivity and consciousness are mutually dependent phenomena, and

---

[29] Both Searle and McGinn can be read as maintaining that there are intentional properties that don't have this feature. According to Searle there are non-conscious intentional states, but they are individuated in terms of their relations to conscious intentional states. McGinn is more sympathetic than Searle to the idea that N-I properties are types of intentional properties.

consequently that we cannot make sense of non-conscious subjectivity or non-subjective consciousness.[30]

If these reflections are correct, then being inwardly-directed is a feature of all and only conscious states. What, then, is the strength of the inwardness intuition about intentional properties? Each intentional property is essentially outwardly-directed. That is what an intentional property is (*being directed thus-and-so*); we individuate intentional properties one from another by reference to the distinct *object* of such directedness. According to the inwardness intuition, some intentional states have the feature of being inwardly-directed, but the inwardness intuition itself fails to specify how and why. Here are three possibilities: (1) Phenomenal properties are essentially inwardly-directed. Intentional properties are not, but when they are "brought to consciousness"—when they are instantiated as part of a conscious state—they acquire the higher-order property of being inwardly-directed. (2) Intentional properties are necessarily inwardly-directed, but not essentially so. The idea might be that the instantiation-conditions for intentional properties require that such properties be subjectively available. (3) Intentional properties are essentially outward-directed *and* inward-directed. Whenever intentional properties are instantiated they present an intentional object *to* a subject—that is what they are.

Option (1) offers no support for Phenomenal Necessity. That intentional properties can be brought to consciousness and thereby become inwardly-directed does not suggest that they *must* be. But if either of Options (2) or (3) are correct, then

---

[30] Some philosophers have diverging intuitions. In 'Against Consciousness Chauvinism,' Itay Shani argues that subjectivity—*having a point of view*—is not a feature exclusively of consciousness. It can be a feature of *agency*. On his notion of agency, a system counts as agential if it has "a complex and globally integrated pattern of organization, one which enables it to monitor how well it is doing in maintaining stability and to behave as a coordinated whole in its ongoing effort to remain stable" (p. 305). Thus Shani thinks there can be non-conscious subjectivity. I suspect (though I am not certain) that Shani's notion of subjectivity is just a different one from that which Searle and McGinn invoke.

Phenomenal Necessity is true—at least if consciousness and subjectivity are connected in the way we have supposed. Is there any way to tell whether one of these ways of cashing out the inwardness intuition is more plausible than the others?

Here are two reasons to think that (1) is not the best way to explain the inwardness intuition. First, there is something strange about the idea of a semantically meaningful item that is not meaningful *for anyone*. It is not the case that *to be is to be perceived*, but perhaps it is the case that *to signify is to be perceived*. Second, it is not clear what it is for the very same property to be instantiated both in consciousness and without. If a property is possibly instantiated by a conscious state, it makes sense to say that it can *only* be instantiated therein.

In sum, here is the case from subjectivity for Phenomenal Necessity: when a subject consciously understands or grasps a content, she does so in virtue of that content's being subjectively presented to her, and such subjective presentation (a) is necessarily connected to consciousness in some fashion and (b) is inexplicable on the basis of functional facts alone. Now, I do not suppose that these ruminations will change many hearts, at least not on their own. For one thing, it is hard to get a firm grip on the subject-matter. (As McGinn says: "We flounder in similes."[31]) For another, the presence of an explanatory gap does not, by itself, have ontological implications. It could be that our concept of *understanding* is bound up with our concepts of *subjectivity* and of *consciousness*, and that there is no a priori connection between these concepts and our concept of a functional system. Yet the existence of such conceptual connections might only imply that our concepts are deficient, rather than that Phenomenal Necessity is true. Nevertheless, it seems to me that the inwardness intuition does provide *some* motivation

---

[31] McGinn (1988), p. 242.

for the Phenomenal Necessity. We might find that its credibility rises if there are other arguments for Phenomenal Necessity that we can more readily evaluate.

5. Phenomenal Necessity: Content Determinacy.

In the second chapter of *Word and Object*, W.V.O. Quine argues that there is never a single, correct way to translate a word. This is because the meaning of a word is a matter of the environmental stimuli that trigger its use, and there are always multiple mappings from environmental stimuli to speaker utterance. If, for example, a linguist wishes to know the meaning of the utterance, 'Gavagai', the fact that there is a reliable correlation between the utterance-type and the proximity of rabbits does not guarantee that the utterance means *rabbit*. It could just as well mean "undetached rabbit parts" or "stage in the life of a rabbit." Further, if the facts about meaning depend entirely on the facts about a speaker's responses to stimuli (as Quine maintains), the upshot of Quine's example is not merely an epistemological point about underdetermination but an ontological point about indeterminacy. The word 'rabbit' in *my* mouth has an indeterminate meaning.[32] Reflecting on these considerations, Searle concludes that Quine has inadvertently constructed a *reductio ad absurdum* of linguistic behaviorism. Because we know from the first-person perspective that 'rabbit' and 'undetached rabbit part' have different meanings, there must be more to meaning than stimulus/response pairs. Searle again invokes his notion of the *aspectuality* of meaning (i.e. its intensionality, its fine-grained-ness[33]):

---

[32] Better: it has a perfectly determinate *disquotational* meaning, and this is all the meaning there is: 'rabbit' refers to *rabbits*.

[33] Searle illustrates the notion by way of non-synonymous, coreferential expressions. I can, for example, think of the selfsame substance under the aspect *being water* and under the aspect *being H2O*. I can think

Behavioral evidence concerning the existence of mental states, including even

evidence concerning the causation of a person's behavior, no matter how

complete, always leaves the aspectual character of intentional states

underdetermined. There will always be an inferential gulf between the evidence

for the presence of the aspect and the aspect itself.[34]

Searle thinks that such an "inferential gulf" exists not just for behaviorist theories of

linguistic meaning, but for any semantic theory that leaves out first-personal intentional

facts: "Any account that uses only a third person objective vocabulary will leave ...

aspectual character underdetermined; because no third person objective vocabulary, by

itself, will be sufficient to characterize all of the aspectual facts."[35] Recently, Horgan &

Graham have reiterated the contention that causal-relational facts about a system are

insufficient to ground determinate intentional content. Take any complex relational fact,

R, that is put forward as the content-grounding relation: "In general, if there is one R-

mapping from a creature's inner states to objects and/or kinds in the creature's

environment, then there are apt to be numerous other such R-mappings as well."[36]

These claims about the gap between intentional facts and third-personal facts

suggest a general argument, from content-determinacy, for Phenomenal Necessity, as

follows: intentional content is at least sometimes both determinate and fine-grained; the

third-personal features of cognitive systems are insufficient to capture or explain these

features; only phenomenality can so explain; hence, consciousness is necessary for

intentionality. In the terminology from the previous chapter: (1) naturalistic theories of

---

of the selfsame object under the aspect *being the Eiffel Tower* and under the aspect *being the tallest iron structure built in France before 1900*.
[34] Searle 1991, p. 54.
[35] Ibid.
[36] Horgan & Graham (2012).

intentionality cannot satisfy *Specificity*; but (2) any adequate theory of intentionality must; (3) phenomenality can; so intentionality must in some sense depend on phenomenality.

The argument is only as good as its first premise, about which there is a rich and complicated literature. (Of course, it is also only as good as the second premise, which I am taking as common ground. And it is only as good as the abductive inference to the conclusion: without a theory of *how* consciousness could be relevant to intentionality, the inference is difficult to evaluate.) So let us look at some particular cases of alleged indeterminacy and the tools with which defenders of naturalistic theories can respond. As before, let 'F' stand for a (sub-propositional) intentional content; let 'O' stand for the relevant class of cognitive system; let '$' stand for some content-bearing structure within a system; let C stand for some causal relation; and let 'A' stand for the type of activation in 'O' that realizes the functional profile of perceptual belief. What is at issue is the claim that, for any C that is put forward as the relation that fixes the content of $, there is some F such that if an F-instance bears C to O, necessarily there is some other property F′ such that an F′-instance also bears C to O. And this despite the fact that O can instantiate states that *determinately mean F*. Examples of problematic pairs of contents (F and F′) come in a variety of types. The first division I want to make is between cases in which the coinstantiation of F and F′ is only nomologically necessary and cases in which their coinstantiation is metaphysically necessary.

So, beginning with alleged instances of merely nomologically necessary co-instantiation: there are at least two important types of case. The first type of case is one where instances of F and F′ lie at different steps in the causal chain specified by C. For

example, if F is the property of *being a rabbit*, F′ might be the property of *being the*

pattern of retinal activation triggered by rabbit-deflected light waves. (Alternatively, F′

might be the property of being whatever the rabbit ate the day before; the property of

being the rabbit's parents; the property of being an explosion that births a cosmos.) I'll

call the problem of satisfying Specificity for such cases "The Stopping Problem." (The

name comes from Strawson [1994]). The second type of case is one where F and F′ are

distinct but coinstantiated properties of a single link on the causal chain specified by C.

For example, if F is the property of being a rabbit, F′ might be the property of *being a*

thing that looks just like a rabbit. Or again, if F is the property being a *rabbit-with-a-*

heart, F′ might be the property of being a *rabbit-with-kidneys*. I'll call the problem of

satisfying Specificity for such cases "The Contingent Coinstantiation Problem."[37]

The advocate of naturalistic theories of intentionality ("the naturalizer"), recall,

has the following resources at her disposal: patterns of causal dependence between

cognitive states of a system and its environment (a lá Causal Covariance Theory);

patterns of causal dependence between cognitive states and the system's homeostatic

needs and phylogenetic evolutionary history (a lá System-Role Theory and Adaptive-

Role theory); and patterns of causal dependence among distinct cognitive states (a lá

Conceptual-Role Theory.) What are the prospects for deploying these resources in

response to The Stopping Problem? As a first step, SRT and/or ART can be used to rule

out properties instantiated by items that are *much more distal* in causal chain C than is F

(i.e. *being a rabbit*). Here is the idea: it is reasonable to assume that there are certain

discriminations that the system needs (or was designed) to make with respect to the

---

[37] Closely related is what Fodor (1990) calls "the disjunction problem"—the problem of resolving what
makes it such that a content-bearing structure denotes F rather than F-or-F′.

environment. Organisms need to track sources of food, sources of danger, etc. True, any state that would be tokened by a rabbit is ipso facto a state that would be tokened by the Big Bang, but viable organisms need to be able to discriminate between distinct effects of the Big Bang. Now, I do not know just how precisely these pragmatic constraints can serve to calibrate a state's mental content—maybe it would work just as well for an organism to track *yesterday's rabbit-dinners* as it would to track *rabbits*—but obviously some progress has been made.

That is as far as SRT and ART can take the naturalizer, however, for the following reason: tracking more proximal causes of rabbits will serve the organism just as well as tracking rabbits. So the naturalizer needs something in addition to biological pragmatics in order to respond to the Stopping Problem. The most frequently proffered solution to the Stopping Problem appeals to a form of counterfactual "triangulation".[38] Suppose that O bears C to an F-instance (O's retinas are stimulated by light refracted off of a rabbit, say), thus causing content-bearing structure $ to enter into content-fixing activation A. $ means *rabbit*, it would seem. But the organism, it is granted, also bears C to F′ (viz., *the pattern of retinal activation triggered by rabbit-deflected light waves*). So does $ mean F-or-F′? Not necessarily, according to the advocate of the counterfactual triangulation strategy. There are many other ways that the rabbit *could have* borne relation C to O, and these other routes would *not* have involved the instantiation of F′. (1) Perhaps O has another sensory modality—smell, say—whereby could have registered the presence of the rabbit. No retinal stimulation would have been involved. (2) Had O encountered the rabbit from a slightly different position, $ would still have been tokened.

---

[38] Dretske (1986), Fodor (2008) and Hill (ms) all provide broadly similar versions of the triangulation strategy.

A *different* retinal stimulation would have been involved. The actual referent of $ is whatever is at the end of all of these counterfactual causings of the tokening of $. So $ means *rabbit*, determinately.

Its thirty years of popularity notwithstanding, the counterfactual triangulation strategy fails rather straightforwardly. There is nothing in the picture so far that rules out disjunctive contents for $—viz., a disjunction of all the rabbitish proximal stimuli that could serve to mediate F's bearing C to O. Triangulation does not help unless there is something that determines that $ has non-disjunctive content—something that fixes the referential "aim" of M onto that non-disjunctive property all the relevant proximal-stimulus states have in common. But what could possibly serve so to determine? I cannot think of anything. Notice: any system that could be used as a reliable rabbit-detection system would also be a reliable rabbitish-stimulus-detection system, because such stimuli are the only way the system could pick up on the presence of rabbits. As far as I can tell, the Stopping Problem remains unsolved.[39]

What about the Contingent Coinstantiation Problem? Again, the natural functions of SRT and/or ART might go some distance. As it happens, whenever *being a rabbit* is instantiated, so is a *being a rabbit-or-a-tiger*. It might make a real difference to viability, however, were an organism to fail to respond differentially to rabbits vs. tigers. Would it make a difference were an organism to fail to discriminate between rabbits and things-

---

[39] Here are two bad alternative solutions:
(1) "Matters of parsimony dictate that we attribute the simplest, non-disjunctive content to the states of a cognitive system." Maybe so, but the Stopping Problem is an ontological problem, not a pragmatic one. Any recourse to pragmatics gives the game away (unless we have reason to think that pragmatic considerations are a good guide to something deeper—but we have no such reason in the present context). (2) "$ would A were Fs to bear C to O' is a law of nature, but disjunctive contents do not enter into laws." I do not know why we should endorse either of these clauses, let alone both together. An sparse, Aristotelian view of laws (as descriptions of the powers of substances) rejects the first clause and an abundant, Humean view of laws (as descriptions of observable patterns) rejects the second—or deflates it into a merely pragmatic principle.

that-look-like rabbits? Likely not: most organisms *do not* so discriminate. And it seems quite obvious that there would not ever be a difference to viability were an organism to fail to discriminate rabbits-with-hearts from rabbits-with-kidneys (as the coinstantiation of the two relevant properties is, to my knowledge, everywhere symmetric).

Jerry Fodor has suggested that there are counterfactuals about content-bearing structures that can adjudicate between rival (coinstantiated) contents.[40] If $ is disposed to A when triggered by a thing that is F, it is, of course, disposed to activate when triggered by a thing that is F′, where F and F′ are coinstantiated. But specify $'s dispositional profile in terms of counterfactual states of affairs in which F and F′ fail to be co-instantiated: $ means F and not F′ just in case, were $ *not* to be triggered by Fs, then neither would it be triggered by *F′s, where the reversal does not hold*. That is, if $'s disposition to be triggered by F′s is "asymmetrically dependent on" its disposition to be triggered by Fs, then $ means F. Applied to our examples: $ means *being a rabbit* and not *being a rabbit lookalike* so long as, were $ *not* triggerable by (mere) rabbit-lookalike-instances, $ would still be triggerable by rabbit-instances, but were M not triggerable by rabbit-instances, its connection with rabbit-lookalike-instances would be lost. Again: $ means *rabbit-with-a-heart* and not *rabbit-with-a-kidney* so long as, were $ not triggerable by kidneyed (non-hearted) rabbits, neither would M be triggerable by hearted (nonkidneyed) rabbits, but...etc.

Does Fodor's appeal to asymmetric dependence succeed? I cannot tell. Note a couple of weird features of the suggestion, as formulated so far. Suppose I were convinced that asymmetric dependence between content-bearing states with different

---

[40] He introduces the idea of asymmetric dependence in Fodor (1987).

types of environmental causes would indeed serve to disambiguate that state's content. And suppose I set out, armed with this useful information, to build a cognitive system whose states contained underived, specific, causal-relationally determined content. I would not be able to do it. Here is why: the system has to *use* some set of proximal stimuli in order to detect distal states of affairs. For any detected property F, there will be a set of proximal stimuli, whereby the system tracks the presence of F. Call the property of being among the proximal stimuli "being the appearance of F." The dispositions of $ to A in the presence of F-instances will be *symmetrically dependent* on the dispositions of $ to A in the presence of instances of *being an appearance of F*. In the actual world, that is, $ is triggered by F's if and only if it is triggered by F-appearances.

Now, there might remain some sense in which F-appearance-triggerings are asymmetrically dependent on F-triggerings. But the asymmetry will not show up *in actuality*; it will only show up in worlds that are sufficiently unlike actuality to break the symmetry. What is unclear to me is whether there is any noncircular way to delineate the possible worlds relevant to disambiguating content. Let me illustrate. Here are two dispositions of mine: (A) raising the temperature of a 68-degree room and (B) being a frustrating interlocutor. There is a clear sense in which disposition B is asymmetrically dependent on disposition A: my being able to have unproductive conversations is dependent on my body's maintaining a viable temperature, plus a bunch of other stuff. But is this so in all possible worlds? Clearly not: were I to find myself in a world with different thermodynamic laws, then the asymmetry goes away. Fodor himself says that content-fixing causal connections have to be law-like. So in worlds in which F does not cause $ to A *are not nomically possible worlds,* and at this point all bets are off. Likely

there will be worlds where asymmetric dependence holds between them in one direction, other worlds where the asymmetry is reversed, and still others—such as the actual world—where the dependence is symmetrical. The point is that unless we have a story about the *grounding* of the relevant counterfactuals, it's impossible to tell whether, and in what sense, content-fixing causal chains and deviant causal chains are dependent on one another.

We do not need to resolve the matter in order to determine whether the indeterminacy objection to causal-relational theories of intentionality can be resisted. That is because there is another class of relevant cases, viz. cases in which the coinstantiation of properties that bear C to O is metaphysically necessary. It was no accident that Quine chose just such a case. We can assimilate it to the present discussion as follows: F and F′ are distinct but *necessarily* coinstantiated properties of an item on the causal chain specified by C. For example, if F is the property of being a rabbit, F′ might be the property of *having undetached rabbit-parts*. (If $ is triggerable by Fs it is also triggerable by F′s because it is metaphysically necessary that a thing is an F only if it is an F′. Hence the two dispositions are symmetrically dependent in all possible worlds; hence Fodor's strategy for dealing with cases of contingent coinstantiation will not apply.) Call this the Necessary Coinstantion Problem. Presumably, if there is a way for the naturalizer to solve this problem, her solution will apply to cases of merely contingent coinstantiation as well.

A solution has been proposed along the following lines.[41] While all rabbits are things with undetached rabbit parts,'There is a rabbit before me' and 'There is an

[41] Cf. Fodor (1994) and Hill (2013).

undetached rabbit part before me' enter into different *inferences*. Suppose that an F-instance bears C to O, result in a content-fixing state as of the perceptual judgment 'There is a $ here'. Suppose also that O believes that (i.e. has a certain beliefwise functional profile to the effect that) it takes at least two parts to make up a whole.[42] If O is disposed to infer "It is not the case that there is exactly one $ before me," then $ doesn't mean F; it must mean something weirder such as F′.[43] The thought is that if you help yourself to all of the inferences that each content-fixing state is disposed to play a role in, you can settle whether that state has the content F or F′.[44] In other words, if you have enough cognitive and logical machinery up and running in the system, you can use the conceptual role of certain content-fixing states to disambiguate their contents, and the problem of indeterminacy is thereby solved or at least mitigated. Call this the "counterfactual inference" strategy.

I argued in the previous chapter that Conceptual Role Theory has no resources to satisfy Evaluability. In brief: unless cognitive systems are infallible at making inferences, not all transitions between content-bearing states are content-determining. Yet the rationality of a transition between two content-bearing states is a function of what content

---

[42] That is, it takes at least two *proper* parts, which is the relevant notion here: the property of being an undetached improper part of a rabbit is equivalent to the property of being a rabbit.

[43] Compare Quine 1960, pp. 52-53: "Point to a rabbit, and you have pointed to a stage of a rabbit, to an integral [i.e. undetached] part of a rabbit, to the rabbit fusion, and to where rabbithood is manifested. Point to an integtral part of a rabbit and you have pointed again to the remaining four sorts of things. Nothing not distinguished in stimulus meaning itself is to be distinguished by pointing, unless the pointing is accompanied by questions of identity and diversity. 'Is this the same gavagai as that?', 'Do we have here on gavagai or two?' Such questioning requires of the linguist a command of the native language far beyond anything that we have as yet seen how to account for." The Fodor-Hill strategy I am discussing is in full agreement with Quine's diagnosis but takes it as an optimistic rather than a hopeless one.

[44] There are two species of indeterminacy that this strategy is hopeless to clear up. Consider disjunctive contents with a necessarily uninstantiated disjunct (such as 'is a rabbit or is a square circle') and conjunctive contents with a universally instantiated conjunct ('is a rabbit and is such that 2+2=4'). A state with the content 'is a rabbit' has exactly the same inferential role as states with these deviant contents. I am disinclined to rest my argument on these monstrosities, though I'm not sure why. (Thanks to Dave Fisher for drawing my attention to this point.)

those states express. In other words, CRT can only make its contributions to the fixing of mental contents *once there are a bunch of determinate contents in place already*; but the counterfactual inference strategy needs the reverse to be true. Let us suppose for the moment that this problem has a solution: that there is something other than content that determines which transitions are irrational, and that the remaining transitions alone determine conceptual role. Here, then is the recipe that the advocate of the counterfactual inference strategy needs to follow in order to respond to the Necessary Coinstantiation Problem:

(1) Distinguish between content-bearing structures that signify properties and those that signify logical functions—i.e., the "predicates" and the "connectives."

(2) Using a functional specification for belief, locate the system's belief-wise semantic constructions; and among these, distinguish between perceptual judgments and other judgments.

(3) Identify the system's dispositions to engage in transitions between belief-wise states. Treat these as inferences.

(4) Match up implicit definitions of the logical functions with the "connectives" whose inferential profile satisfies those definitions.

(5) With the interpretation of the "connectives" in hand, look back at the transitions the systems makes from its perceptual judgments; use these inferences to determine its judgments about sameness and difference with respect to objects in the environment; use these judgments to determine the extension of the "predicates."

So runs the last, best hope of the naturalizer for responding to the Necessary Coinstantiation Problem. (If it works it will also serve as a solution to the Contingent Coinstantiation Problem). Lesson 1: causal-relational theories of intentionality have more resources at their disposal for dealing with indeterminacy worries than the likes of Searle and Horgan & Graham would let on. Lesson 2: a lot has to be in place for these resources to gain any traction. As we have seen, to implement (3), the naturalizer has to be able to distinguish the rational from the irrational transitions, without invoking content. Moreover, the naturalizer has to be able to distinguish the rational from the *a*rational transitions, i.e. mere associative causings of one occurent belief by another. (My perceptual judgment "There exists a frog" might occasion the occurrent judgment "Frogs are peculiar" despite the fact that I'm not inferring the latter from the former.) To implement (2), the naturalizer needs a functional specification of belief—a theoretical component to which many gesture but which nobody has actually secured.[45] To implement (1), the naturalizer needs to be able to be able to syntactically carve up the cognitive system—which means that she needs to be able to individuate the system as a unified cognitive system. One challenge to this task is that the system could expand and change over time. (Note the ensuing complications for step (4). Are the meanings of the "connectives" determined by the system's inferential dispositions *at a time* or *across time*?) Another challenge anticipates the final argument for Phenomenal Necessity we

---

[45] Here are a couple of reasons to doubt that there could be an adequate functional specification of belief, at least assuming that such an account will make mention of a system's behavioral outputs. (A) Beliefs (partly) cause actions, according to current orthodoxy in philosophy of action. Actions admit of many descriptions, some of which capture the essence of what the agent is doing and some which only describe accidental features—and the best (the only?) way to tell the difference is by recourse to their causal ancestry. In other words, actions are individuated in terms of the beliefs that cause them; so beliefs can't be individuated in terms of the actions they cause. (B) A functional specification of belief rules out the possibility of inanimate believers such as Strawson's (1994) "Weather Watchers". It would be odd to rule out the possibility of such beings from the armchair.

will be examining: is there a deep fact of the matter about which states of the world count as part of a single cognitive system, and which do not?

Summing up: it has been claimed by Searle and like-minded thinkers that consciousness is necessary for intentionality because the only potential determiners of the contents of a system's mental states are (a) the causal-relational facts about the system and (b) the phenomenal facts about the system, but the most sophisticated appeals to causal-relational facts fail to explain the degree of determinacy that we know our mental states to have. Specific challenges come in the form of the Stopping Problem, the Contingent Coinstantiation Problem, and the Necessarily Coinstantiation Problem. The first remains unsolved. Attempts to solve the second are hard to evaluate. To solve the third, naturalizers have proposed the counterfactual inference strategy, the success of which depends on the possibility of characterizing certain features of a cognitive system *without recourse to the system's semantics*, including (1) the distinction between rational and irrational transitions; (2) the grammatical type of its content-bearing items; (3) those functional states which amount to occurrent beliefs; and, most foundationally, (4) the system's individuation-conditions. I turn now to arguments for Phenomenal Necessity that turn on pessimism about (4).

6. Phenomenal Necessity: Cognitive Unity.

In the present section I explore a number of related suggestions to the effect that only consciousness can serve to *individuate* or *unify* mental phenomena. I again begin with some intriguing but underdeveloped comments by Searle.

> There are deep reasons having to do with the nature of mental phenomena whereby our notion of an unconscious mental state is parasitic on our notion of a conscious state.[46]

> So here are two states in me, my belief and my axon myelination; both have something to do with my brain; and both are unconscious. But only one is mental, and we need to get clear about what makes it mental and the connection between that feature—whatever it is—and consciousness.[47]

Searle is gesturing at two senses in which consciousness is necessary for the individuation of mental phenomena. The first sense is conceptual: our *notion* of an unconscious mental state is parasitic on our notion of a conscious state. The idea, I gather, is that any stable conception of the mental is anchored in a conception of the phenomenal. The point is interesting; if Searle is right, then any scientific study of the mind has to delimit its subject matter in terms of consciousness. Now, it does not straightforwardly follow that consciousness would need to show up in a scientific *theory* of the mind, any more than the surface features of water need to show up in a chemical theory of H_2O. Nor does it straightforwardly follow that it would not, of course. It would be a fruitful philosophical task to evaluate and explore the implications of Searle's claim that our notion of consciousness is conceptually prior to any other mental notions, but I will not pursue that task here.[48]

---

[46] Searle (1991), p. 49.
[47] Ibid., p. 50.
[48] In chapter 1 of *The Sources of Intentionality*, Kriegel develops an elaborate argument to the effect that we derive our conception of intentionality from our awareness of the intentional features of our conscious experiences, rather than from some third-personal source. I find the argument compelling, but again, it is not clear just what is supposed to follow. Conceptual priority is not, eo ipso, ontological priority. (And this notwithstanding the fact that Kriegel titles his chapter "The Experiential Origins of Intentionality".)

According to the second line of argument suggested by the above passages, one must invoke consciousness in order to separate the mental from the non-mental features of a cognitive system. The claim requires (a) that nothing but consciousness can do the trick, and (b) that consciousness actually can. Searle accounts for these requirements by appealing to his indeterminacy-based arguments for Phenomenal Necessity. All and only mental states have "aspectual shape," he argues, and only consciousness can account for aspectual shape. As Ludwig (1993) notes, Searle's strategy requires that all mental states have aspectual shape (i.e. fine-grained intentional content), but such a requirement is controversial: feelings and emotions arguably need not, and there is debate even about sensory states such as the experience of after-images and so-called "ganzfelds" (i.e. phenomenal fields of uniform color). Is there a less contentious defense of the claim that only consciousness can serve to sort mental from non-mental phenomena (i.e., one that doesn't rely on *determinate content* to do the sorting)?

In order to make progress, it will be helpful to distinguish between three distinct explanatory questions related to cognitive unity. For some cognitive system O,

CU1: What unites the mental states of O as states of one mind rather than of several?

CU2: What unites the mental states of O as mental states rather than as non-mental states?

CU3: What unites O as a single cognitive system rather than as several systems or as part of a system?

These are important questions for a variety of reasons. For one thing, if there is no correct answer to CU1, we cannot make sense of the discipline of epistemology—at least insofar

as it studies the dynamics of rationality and justification. Epistemologists are interested in what it is *for a subject* to be rational, to be justified in believing p, etc. If there is no answer to the question, "Whose beliefs are these?", then there cannot be an answer their epistemic credentials. I assume that such a result would be unacceptable.

Ludwig has argued that only consciousness can provide an answer to CU1.[49] His argument amounts to a process of explanatory elimination: either a state is part of a single mind (a) because it bears some relation (identity, causality, etc.) to a subject's conscious states, (b) because it is causally related to a subject's body, or (c) just because (i.e. the explanatory relation is brute). The relation cannot be brute, lest it be possible that any state of any body could be paired with any subject at all—or, at any rate, lest the impossibility of weird state-subject pairings be inexplicable. But a state's biological-functional connection with the subject's body will not work either:

> The relations that hold between a person and his body that makes it his are that changes in it affect his mental states, and in particular his conscious mental states, and that his beliefs and desires explain its behavior, and more generally, his mental states affect more or less immediately his body. But it seems clear that two different people could bear these relations to one body, either at the same or different times, as is shown by the possibility of conceptualizing cases of multiple personalities as cases of multiple persons occupying a single body.[50]

I think Ludwig is right that the "same biological system" criterion for being states of one mind will not work, but not for the reason he gives. I suspect that the naturalizer will respond to Ludwig as follows: a subject does not stand in a causal relation to her cognitive states; rather, a subject *just is* a causally integrated network of cognitive states. So while it is possible for a state of a cognitive system to be causally affected by some other subject—say by hooking up one brain to another via radio link—mere causal

---

[49] In Ludwig (1996), pp. 20-24.
[50] Ibid. p. 23.

influence on that state wouldn't be enough to disrupt its causal *embeddedness* within a single mind. Now, causal embeddedness is a degreed notion, but that's no matter, since, plausibly, so is cognitive integration. If, for example, a cognitive system were typically inclined to act as though P and occasionally to act as thought not-P, there might be no stable sense in the system believes P. More dramatically, if the functional mapping of a brain seemed to settle into two largely independent but massively causally integrated subsystems, we might be inclined to treat that brain as housing two minds. (There might be good reasons to reject a causal-integration account of subjecthood, but nothing in Ludwig's argument provides such reasons.)

But so saying just pushes matters back onto an answer to CU2. States of brains are causally embedded within the functioning of an entire organism. What makes it the case that some of those states are mental and some are not? Here, again, there is a plausible naturalistic response (that eschews phenomenology): cognitive states are those that play a role in cognitive processes, and cognitive processes are those that explain certain sorts of *behavior*. Which sorts? Well, we might appeal to the way that cognitive scientists discriminate cognitive vs. non-cognitive behavior. The consensus seems to be that cognitive behaviors are those that exhibit certain forms of "flexibility." Paradigms of cognitive flexibility include responding to stimuli differently in different environments; responding differentially to superficially distinct items that nevertheless fall within the same metaphysically deep category; being able to coordinate information within distinct sensory modes (visual vs. auditory, say); and so on.[51] The general idea is that if an organism exhibits cognitively flexible behaviors, then the states that produce those behaviors count as cognitive. Now, I am not *sure* that this criterion will be good enough;

---

[51] See Buckner (2013) on the notion of flexibility in cognitive science.

after all, a whole lot of organismic functioning is relevant to the production of its

behavior, cognitive or otherwise. But if there are physiological processes that are

implicated *only* or *predominantly* in the production of cognitive behavior, perhaps the

criterion *will* be good enough. (Note: I suspect that cognitive scientists treat cognitively

flexible behaviors as *indicative* rather than *constitutive* of cognition. If I am right, it is a

major liability that the naturalizer fails to endorse the order of explanation espoused by

scientists—a point to which I return presently.)

This answer to CU2 presumes that an answer to CU3 is available: in order to

determine whether a system exhibits cognitive behavior, we need a target of our

investigation, a more or less delineated entity whose behavior we're examining. To see

why this is a problem, we need a better grip on what a *system* is. I take it that systems are

entities (or aggregates) that exhibit functional structure with at least the following two

features: *interdependence* and *complexity*. An interdependent system is one whose

functional structure includes feedback loops. A complex system is—well, a complex one,

where complexity could come to *quantity* of processes (e.g., the number of processes

running in parallel) or *quality* of processes (e.g., the way that processes iterate or embed).

Obviously, interdependence and complexity come in degrees.

Answering CU3 is hard for three reasons. First, the world has lots of systems.

River eddies exhibit complexity and interdependence. So do cells. So do star clusters.

How are the *cognitive* systems to be distinguished from the rest? Call this *the profusion*

problem. Second, systems are not functionally isolated from their environments.

Organisms are causally embedded in ecosystems on which they are thoroughly

dependent. From the "perspective" (so to speak) of a carbon molecule, there is no grand

difference between being part of an organism, part of the soil, or part of the atmosphere. Further, systems *nestle*: cells are parts of tissues which are parts of organs which are parts of physiological systems which are part of organisms which are parts of ecosystems which are parts of the biosphere. (And that's just the nestling that occurs among biological systems.) How are boundaries around individual systems to be drawn? Call this *the boundary problem.* Third, systems can overlap. A river's ecosystem can cut right through—and consequently supply a part of—a mountain's ecosystem and a valley's ecosystem; a roaring fire can serve both my home's heating-system and its system of illumination. Call this *the overlapping problem*. I take it that any viable answer to UC3 has to be able to solve all three of these problems. I do not think any can, though. Here are some attempts:

(1) The non-answer: "Biologists study organisms. The category of an organism is thus in good standing. Cognitive systems are the organism-like systems." Obviously the category is in good *methodological* standing, but such a status fails to guarantee that it is in good metaphysical standing. (Here we find vindication for Ludwig's rejection of a biological criterion for system-individuation.) I take it that biologists typically *presuppose* that the category of an organism is in good metaphysical standing. The question of what explains that good standing remains open. Further, it is an epistemically open possibility that biological systems are not particularly well individuated after all, the success of both folk and scientific biology notwithstanding. (I take it that *species* has turned out to be a metaphysically shallow category. Or again: in a Dawkinsian mood someone might conclude that the category of a gene is much deeper metaphysically than the category of an organism.) The fact that biologists are unthreatened by the

metaphysical question of organism-individuation does not mean that they have or need an answer, or even that there is one.[52]

(2) The pragmatic answer: "Any system can be treated as cognitive, depending on our theoretical interests." Every event in the history of the world is minimally embedded in the world's causal structure. So every event can be treated as cognitive. On a realist reading of the pragmatic answer, that is because every event *is* cognitive—which I take to be an unacceptable result. On an anti-realist reading, an event's being cognitive is a matter of our treating it in certain ways. It is not clear that such cognitive instrumentalism is so much as coherent (as many philosophers have pointed out): presumably, *treating* a event as cognitive is itself a cognitive event—which, according to the present view, is cognitive only in virtue of some other treating-as-cognitive event. In general, there is no such event (nobody is treating-as-cognitive the thoughts of the cognitive scientists while they work), and even if there were, it would only amount to the next step in a regress. Even setting aside this worry about coherence, the anti-realist reading is implausible on its face: we know via acquaintance with the cognitive features of our own conscious mental states that cognitive phenomena are objectively real features of those states.

(3) The criterial answer: "Cognitive systems are those systems that admit of interdependence and complexity above threshold such-and-such." The first problem is that any way of providing a cut-off is going to feel arbitrary. Does a system with five causal nodes and two nestled feedback loops count as cognitive? With five hundred nodes? Five million? Why? It is hard to imagine answers to such questions, and

---

[52] Compare Strawson (2008): "Some philosophers ... may now say that there is something very special about the representational states of biological entities ... because their capacities to enter into such states were 'designed' by evolution or specific tasks. But there is no deep difference here, in the great story of the Universe" (p. 287).

consequently it is hard to imagine a criterial response to the abundance problem.[53] I

suspect that the boundary problem and the overlapping problem pose insurmountable

challenges to the criterial answer, owing to the fact that systems that satisfy the stated

criteria could have shared parts. Take some system, O, that is nestled within system O+,

that itself embeds system O-, and that overlaps with system O′. Suppose there is some

structure $ that is a part of all such systems. Now recall, from the discussion of the

counterfactual inference strategy above, that a naturalizer's best hope to ground

determinate mental content in a cognitive system is to appeal to the inferential role of a

system. All of this adds up to the possibility that $ plays a different role in all four

systems and consequently has four different representational contents. That is: a state's

having a content is not only a relational feature of that state; it is a feature that that state

has *relative to a system*, and it could have many—perhaps an infinite number—of such

features, depending on how many systems it is a part of. Now, might a conscious mental

state of mine be one such state—a state that is a part of many systems with a sufficient

degree of interdependence and complexity? I fail to see why not. But it seems obvious

that my mental states have the (unique) contents they have and none other. "No, they

have unique contents *from your perspective*, i.e. from the perspective of the cognitive

system that constitutes your subjective standpoint." I have no knock-down argument

against that suggestion. Note that brains are excellent candidates for housing profusions

---

[53] Itay Shani (2008) thinks that only certain types of complex interdependence are sufficient for *autonomy*, which provides a criterion for being a cognitive system: "while systems like hurricanes and vortices manifest processes that contribute to their own far-from equilibrium stability, they do not possess a genuine capacity for action. What a genuine AA [autonomous agent] possesses, and a lesser self-maintaining system lacks, is *a complex and globally integrated pattern of organization*, one which enables it to *monitor* how well it is doing in maintaining stability and to behave as a *coordinated whole* in its ongoing effort to remain stable" (p. 305). So long as we keep in mind that Shani uses such terms as "monitor" and "behave" and "effort" in entirely impersonal terms (on pain of explanatory circularity), I see no reason to doubt that a hurricane could implement the type of complex feedback system Shani has in mind.

of nestled and overlapping systems. There is a distinct possibility, on the criterial answer, that my mental states are shared by many cognitive systems. I find the suggestion both bizarre and disconcerting.

We can combine the discussions of the present section with the previous section (on content determinacy) to construct the shape that a complete naturalistic theory of cognition would have to take.

Because some cluster of phenomena is sufficiently causally interdependent and complex, it counts as a cognitive system; because certain of its processes are implicated in behavioral outputs of a special sort, the states that drive such processes count as cognitive. Because of their dispositions to engage in certain transitions (per the presuppositions of the counterfactual inference strategy), a subset of these states count as expressing logical functions; because of the way these "logical connectives" team up with another class of cognitive structures, "the predicates" (and because of the way "the predicates" track environmental stimuli), the whole system counts as having particular mental states with determinate content.

Such a story strikes me as explanatorily backward. For example, it seems obvious to me that cognition is metaphysically prior to intentional action, not the other way around. More generally: it fails to take cognition seriously as a phenomenon that carves nature at the joints, rather than as an epiphenomenon that occurs only relative to arbitrarily delineated chunks of causal reality.[54] If the naturalizer has to go in for such a story, that is a good reason to reject naturalism and instead appeal to consciousness in accounting for

---

[54] Writes Galen Strawson: "No doubt other reasons could be given for discerning UNA ["underived aboutness in a non-experiential entity"] only at certain special points in the great nexus of cause and effect. I think, though, that they are bound to be metaphysically superficial." I agree.

cognitive unity. Here is how the alternative, consciousness-based method of individuation would go: what unites the mental states of O as states of one mind rather than of several (per CU1) is that they are actual and potential conscious states of the selfsame subject; what unites the mental states of O as mental states rather than as non-mental states (per CU2) has to do with their being actually or potentially conscious; and what unites O as a single cognitive system rather than as several systems or as part of a system (per CU3) has to do with the system's giving rise to a single stream of consciousness. In sum: on this consciousness-based theory of cognitive unity, there is a natural order of explanation that *begins* with conscious mental states and *ends* with the individuation of systems that support cognition, and what makes it the case that such systems engage in "cognitive behavior" is that mental phenomena play a central role in their causal ancestry.

7. Conclusion.

In this chapter I have reviewed arguments to the effect that consciousness and content have a strong connection to one another. I first reviewed arguments for the claim that some intentional properties supervene on phenomenal properties (a claim I called "Phenomenal Sufficiency"). Arguments for Phenomenal Sufficiency take their cue from three different phenomenological intuitions: (1) we cannot conceive of phenomenal duplicates who fail to share intentional properties; (2) we cannot conceive of certain types of intentional contrasts that fail to be accompanied by corresponding phenomenal contrasts; and (3) phenomenal states afford us with discriminatory abilities with respect to the intentional properties of those states. I concluded that these arguments give us strong reason to think that phenomenal properties in perceptual and imaginative states—

and perhaps also in purely cognitive states—are metaphysically sufficient for those states having certain forms of content.

I then reviewed arguments for the claim that some phenomenal properties supervene on intentional properties, a claim I called "Phenomenal Necessity." According to one argument, when a subject instantiates an intentional property in such a way that she *grasps* its content, such an intentional property must be inwardly-directed toward the subject (while being outwardly-directed toward an intentional object), and consciousness (and hence the instantiation of phenomenal properties) is required for such inward-directedness. According to a second argument, intentional properties present contents to the subject that are determinate in ways that no naturalistic theory can capture, but that phenomenalist theories presumably can capture—so unless there are other theoretical options (besides naturalistic and phenomenalist theories), subjects who can entertain determinate intentional contents must at least be capable of instantiating phenomenal properties. According to a third argument, intentional properties clump together, in isolated pockets of reality, as properties of unified cognitive systems. But no naturalistic theory can capture such unity, whereas phenomenalist theories presumably can—so unless there are theoretical options besides these two, cognitive systems must at least be capable of instantiating phenomenal properties. I concluded that, while the first argument is hard to frame in a rigorous way and hence difficult to evaluate, the second and third arguments can be framed rigorously and forcefully—giving us strong reason to think that a *capacity* for instantiating phenomenal properties is metaphysically necessary for a system to instantiate any intentional properties.

Perhaps there are other reasons, besides those canvassed here, that have convinced philosophers of a modally strong connection between consciousness and content. I believe I have touched on all of the most influential reasons, however. None of these reasons has been both fully developed in the literature and favorably received by all parties. I hope that the foregoing has shown that, as these reasons are developed into rigorous arguments, they appear more formidable rather than less. Some form of the Inseparability Thesis is very likely true.

3. A Theory of Phenomenal Grounding

1. Introduction

In the last chapter I discussed versions of the "Inseparability Thesis," according to which, for some class of intentional properties and some class of phenomenal properties, a subject instantiates items in the first class iff she instantiates items in the second class. I discussed the Inseparability Thesis, and the arguments that count in its favor, in the endeavor to motivate a stronger claim, viz., the Phenomenal Grounding Thesis: all instances of intentional properties are partly grounded in instances of phenomenal properties. But there are two logical gaps between the Inseparability Thesis and PGT. First, the Inseparability Thesis is a mere biconditional, whereas PGT posits an asymmetric dependence-relation. Why should we think that phenomenal properties *ground* intentional properties, rather than the other way? Second, the Inseparability Thesis is existentially quantified, whereas PGT is universally quantified. Why think that *all* intentional properties are grounded in phenomenal properties?

There are two reasons to think that the phenomenal properties ground intentional properties and not the other way around. First, there are some phenomenal properties whose instantiations have no implications regarding intentionality: some somatic and conative experiences, e.g., feelings of distress or exhilaration, lack intentional content. Whether sensory experiences ever lack intentional content is less clear, but it does seem to be the case that the same intentional content can be presented in distinct sensory modes: I can seem simultaneously to *see* my pet owl and to *smell* my pet owl. Further, a single sensory experience can present the same content in different ways (e.g. visually

97

blurrily or visually clearly).[1] Now, philosophers who pursue an intentionality-first

explanation of the Inseparability Thesis—e.g., reductive representationalists—are not

likely to just give up at this point. According to the standard representationalist story,

somatic and conative experiences *do* have intentional contents after all, and the

phenomenal differences between intentionally-equivalent states can be accounted for in

terms of the functional role of those states.[2] There is no space here to review and evaluate

the details of these proposals. For present purposes, I simply report my own experience of

incredulity at such strategies. I cannot understand the suggestion that all there is to a

feeling of pleasure or to the visual experience of a rainbow is *representational content*

plus functional role.

A second reason to think that intentionality depends on phenomenology is that the

intentionality-first proposal cannot capture the intrinsicality of phenomenology. In

chapter 1 we explored prospects for a naturalistic theories of intentionality. It turns out

that N-I properties—the properties with which naturalistic theories of intentionality

identify the intentional properties of our acquaintance—are not the sort of properties that

can supervene on the intrinsic properties of occurrent mental states. But that means that

they are bad candidates to serve as the reduction-base for phenomenal properties: if we

can be introspectively acquainted with phenomenal properties, then those properties must

be intrinsic to the states that include them.[3] But it is hard to come up with any alternative

to N-I properties, if we want to account for intentional properties in a non-

phenomenalistic way. In short: intentionality-first explanations of the Inseparability

---

[1] See Loar (2002a and 2003b) for more detailed discussion of these points.
[2] Harman (1990), Dretske (1995), and Tye (1995) are to be credited with getting these proposals up and running.
[3] See chapter 1 section 3, as well as chapter 2 footnote 2.

Thesis entail that phenomenal properties are not intrinsic properties of occurrent mental states, and that's an unacceptable result.

Whence the leap from an existentially-quantified Inseparability Thesis to a universally-quantified Phenomenal Grounding Thesis? Here the inference is an abductive one: I submit that PGT is the best explanation of the Inseparability Thesis. Whether I am right about this can only be determined in light of the details of a *theory* of the phenomenal grounding of intentionality. So let us turn to our attention to the shape of such a theory.

2. The Nature of P-I Properties.

Here is a sketch of my theory of phenomenal grounding. Among the phenomenal properties are *phenomenal-intentional properties*, or "P-I properties". The nature of a P-I property consists in the presentation to the subject of an intentional object. An example is the P-I property whose intentional object is *causation*. When a subject instantiates this property, she is thereby consciously presented with *causation*. If a subject attends to a phenomenal state that includes this property, she can thereby be acquainted with it; and because its nature consists in the presentation of *causation*, her introspective acquaintance with a state that includes this property affords acquaintance with the intentional object of this property, *causation*. Hence she can be acquainted with paradigmatic intentional contents of her mental states (per the Acquaintance Argument of chapter 1).

What is the relationship between P-I properties and other phenomenal properties? One possible answer is that P-I properties are *reducible* to other phenomenal properties.

Another possible answer is that they form a *sui generis* type and are thus irreducible to other phenomenal properties. I take it that the former option is preferable, all else being equal, because more parsimonious—it does not require expanding our inventory of the basic phenomenal properties. The reductionist project has recently been explored by Katalin Farkas and Farid Masrour. I will explain their approach, why I think it fails, and why we should expect a similar fate to befall future attempts at reductionism as well.

Katalin Farkas (2013) observes that the difference between phenomenal episodes that seem to present an objective world and those that do not runs broadly parallel with the difference between phenomenal episodes that admit of certain systematic correlations between somatosensory properties and those that do not. Somatosensory properties, or S-S properties, comprise visual, auditory, tactile, proprioceptive and affective sensations—in short, all the phenomenal properties that are not intrinsically intentional. Farkas points out that S-S properties such as auditory and visual sensations are often systematically correlated: sounds grow fainter as our visual image of the sound-source grows smaller. Further, experiences of these sorts systematically correlate with agential and proprioceptive experiences related to bodily movement, as we move and act and investigate. In contrast, pain-sensations and visual sensations such as after-images are not subject to these same sorts of inter-modal correlations, and neither do they seem to present an objective world.[4] Noting this close connection between phenomenal covariance on the one hand and phenomenally-grounded intentionality on the other, Farkas proceeds to *reduce* the latter to the former:

---

[4] Writes Farkas (p. 109): "a difference between ordinary perceptual experiences on the one hand, and experiences that don't seem to present experience-independent objects—afterimages, phosphemes—on the other...[is] that the simpler phenomenal features of perceptual experiences are organized into a systematic, cross-modally coherent and predictable order. This order is what I call the 'structure' of the experience."

> A feeling may just be a feeling and not present or represent anything. However, when these sensory features are received by the subject in a highly organized and predictable structure...the experience may become suggestive of the presence of something *beyond* this experience, namely, an experience-independent object. Perceptual intentionality is thus constituted by the structure of sensory phenomenal features: by the way these features hang together and respond to movement and inquiry.[5]

The idea, I gather, is that sensory experiences seems to present mind-independent states of affairs when and only when S-S properties, both intra-modally and inter-modally, covary in certain ways, and hence that apparent mind-independence ought to be explained in terms of such covariance. Farkas is obviously on to something: we frequently exploit the correlation between phenomenal covariance and phenomenal intentionality in order to resolve ambiguities in what our perceptual sensations seem to present to us. For example, suppose you have an auditory sensation of a high pitch and you want to know whether your ears are ringing or whether you're hearing a noise: a normal course of action would be to walk to a different room and attend to whether the sensation changes.

I don't think that Farkas's proposal will get us very far on its own. Depending on what sorts of experiential features can contribute to the relevant patterns, the proposal either under-predicts or over-predicts intentional contents. Consider a case of fairly static phenomenal goings-on: waking up in the morning and staring, unmovingly, at your white ceiling. In such a case you supply very little "movement and inquiry" for your visual phenomenology to covary with. By hypothesis, your experience does present intentional contents to you (you seem to see a white surface at some distance from you), but because your experience is so sparse and static, it is hard to locate much phenomenal covariance within it. In response, Farkas could say that your experience exhibits covariance between

---

[5] Farkas (2013), p. 100.

absences—viz., between your absence of movement, on the one hand, and an absence of a change in visual phenomenology, on the other. But if absences can contribute to the relevant patterns, all experiential episodes will be too richly patterned to explain where intentional content comes from. Add to our example that your first few moments upon waking include a steady headache. The covariance between your absence of movement and the absence of a change in head-pain does not make it seem to you that there is an object external to the mind that is painful.

So, something needs to be added. One possibility would be to let phenomenal patterns accrue over one's lifespan. While this may go some distance in helping explain the ceiling-staring case—perhaps your ceiling-caused visual phenomenology *does* covary with movement and inquiry, if we are taking into consideration all past ceiling-caused visual experiences—it is going to run into standard puzzles about induction. What about patterns involving not only *absences* of phenomenal properties but also *disjunctions* of phenomenal properties? Do unprecedented pairings of phenomenal properties amount to revisions of old patterns or introductions of new ones? Probably some additional element is needed to constrain the *types* of patterns that are relevant to determining content. A proposal very similar to Farkas's but that seems to build in this extra element from the start is that of Farid Masrour (2013). He says that "schematic dynamical unity"—his analog to Farkas' "structure", and which he appropriates from his reading of Kant—is constrained by the deployment of what he calls "schematic representations" within phenomenal states:

> It helps to distinguish between the Kantian thesis and a simple regularity account of phenomenal objectivity. Such an account identifies phenomenal objectivity with regularity in the course of experiences. On this account it suffices that representations of some properties co-vary in accordance with some rule. But it is

not required that these rules be associated with a specific form of phenomenology nor that they depend on a specific psychological mechanism. In the Kantian view, in contrast, the experience of schematic dynamical unity requires the activation of schematic representations which are associated with a specific form of phenomenology. A schematic representation is in effect a detector whose job is to test whether certain dynamic relations between the values of certain representations obtain. An abstract arbitrary rule that does not correspond to such a detector does not count as a schema.[6]

The general shape of Masrour's view (as I read him) is as follows. Not only are our sensations systematically correlated in certain ways. We also have a phenomenological awareness of certain of these correlations. This awareness comes in the form of what Masrour calls "implicit anticipatory conditionals"—i.e., a sense of what's expected and what's not with respect to the unfolding of our perceptual experiences. When our experiences conform to these implicit anticipatory conditionals, and *because* they do so, they seem to present us with an objective world. The right sort of phenomenal awareness of the right sort of phenomenal patterns determines intentional content.

Attempts such as these to provide a reduction of phenomenally-grounded intentionality to non-intentional phenomenology are intriguing and worth developing. Nevertheless, such reductionism—at least of the sort developed by Farkas and Masrour— is subject to two serious challenges. The first has to do with the fact that phenomenally-grounded intentionality can be found outside of the realm of the perceptual, in imagination (less contentiously) and in cognition (more contentiously). I take it as data (a) that our imaginative episodes present intentional contents, and (b) that they do not admit of the sorts of inter-modal correlations found in perception. (Indeed, types of phenomenology that are correlated in perception can be willfully de-coupled within our imaginative flights of fancy.) That there is purely cognitive phenomenology is less

---

[6] Masrour (2013), p. 132.

evident, but if there is, the possibilities for semantic recombinability of thought-constituents guarantees that cognitive-phenomenal covariance will be scarce. It is open to the reductionist to deny that the content of conscious thoughts is determined by the phenomenology of conscious thoughts, but rather is non-conscious or merely latent or some such.[7] I do not think the reductionist is so easily off the hook regarding imagination, however: it just isn't plausible that what imaginative episodes are like phenomenally and what they are like intentionally are separable. When we imagine, it is not as though an undifferentiated blur of colors and shapes comes before our minds that we subsequently *interpret* or *recognize*.[8]

Now, Farkas and Masrour could respond by saying that they never meant their theory to reduce *all* phenomenally-grounded intentionality. So long as their theory covers phenomenally-grounded intentionality *in perception*, then they will have succeeded according to their aims, recognizing that imagination calls for a different theory.[9] But it strikes me as implausible that the phenomenally-grounded content of perceptual states and the phenomenally-grounded content of imaginative states are of two different metaphysical types. At any rate, given our ability to tell when a perceptual state and an imaginative state have the same or different content (see chapter 1), it would be ad hoc to posit a deep metaphysical difference between the two types of content, in the absence of strong theoretical reasons to do so.

The most promising route for the reductionist would be to try to treat imaginative content as somehow parasitic on or an extension of perceptual content. The idea might be

---

[7] I myself develop a fairly austere view of cognitive phenomenology. See chapter 5, section 6.
[8] McGinn (2005) argues this point at some length.
[9] Farkas grants that her theory does not apply to non-perceptual phenomenology and hence that we will need a disjunctive theory of phenomenally-grounded intentionality. This is where I *tollens*.

something like this: when I imagine something—a red ball, say—what makes it the case that my sensations present red-ball-hood to me is that I treat those sensations as having the same covariance-profile that they have in perception. That is, when I have had reddish-roundish sensations that have enjoyed a certain covariance-profile in connection with other sensations, I come to have certain anticipations with regard to the reddish-roundish sensations. Seeming to perceive a red ball *just is* having this set of anticipations toward a reddish-roundish sensation. I can subsequently *imagine* a red ball by (a) imagining a reddish-roundish sensation, and then (b) imagining having the set of anticipations toward it. I don't think this makes any sense, though; can't I imagine having a different set of anticipations toward a red ball—i.e. that it will suddenly change color and shape and so forth—while still imagining a red ball? Perhaps there is way to develop the idea that I haven't considered, but I am doubtful.

The second challenge has to do with the relevant determination-relation. Farkas and Masrour gesture at two different accounts of it. They both maintain that the obtaining of certain patterns within one's experience (Farkas), or the *awareness* of those patterns (Masrour) *constitutes* phenomenal intentionality. (I'll stick with Farkas's simpler formulation henceforth.) But elsewhere they describe the relation in epistemological terms. I have quoted Farkas as saying that when experience exhibits phenomenal covariance, it "may become *suggestive* of the presence of something beyond" it (italics mine). And I have quoted Masrour as saying that a schematic representation is "a detector whose job is to *test* whether certain dynamic relations between the values of certain representations obtain" (italics mine). These formulations suggest that phenomenal covariance *justifies* a subject in treating an experiential state as intentional. But there are

problems with both the metaphysical and the epistemological renderings of reductionism: both turn out to face the same dilemma.

Suppose we start from the epistemological rendering. The idea here would be that phenomenal covariance serves as the evidential input into an abductive inference, the output to which would be that such-and-such a content is perceptually presented to one. The trouble with this idea is that evidence, and judgments made on the basis of that evidence, are *distinct existences*. Phenomenal covariance does not by itself ground phenomenal intentionality, on the epistemological rendering, in other words; in addition, an act of perceptual interpretation is required.[10] Now, is this act itself conscious? If not, the reductionist has dispensed with the Perceptual Grounding Thesis altogether. But if it is conscious, then the reductionist has appealed to an unreduced instance of conscious intentionality, and her theory is incomplete.

We probably ought to take the epistemological language that Farkas and Masrour use as loose or metaphorical, and opt for the metaphysical rendering of reductionism. But the metaphysical rendering runs into trouble, too. On this version, the phenomenal patterns exhibited by a state stand in the constitution-relation to the phenomenally-grounded intentionality of that state. But it is not clear what it is for phenomenal patterns to *constitute* intentionality. There are at least two explanatory questions that come up for the proposal: (1) Why would instances of phenomenal covariance constitute intentional content at all? (2) Supposing they do, what makes it the case that they constitute particular contents rather than others?

---

[10] I am not using "act" here in the sense of *action*: I'm not suggesting that the agent would need to engage in perceptual interpretation intentionally or voluntarily or anything like that. Take "act of perceptual interpretion" to mean *mental episode that supplies the subject with an interpretation of a perceptual state*.

As regards (1), compare: the pattern of brushstrokes in a painting constitutes the painted image. And the painted image depicts something (e.g. a pond with lilies). But of course the patterns do not, in themselves, constitute anything *intentional*. (If I recognize the painting as a picture of a pond, that is because I, the viewer, visually interpret it as such.) So it is hard to make sense of the claim that phenomenal patterns constitute intentional content. There are interesting similarities between this claim (about the constitution of intentional content by phenomenal patterns) and the claim made by advocates of Conceptual Role Theory to the effect that a functional system can come to have content in virtue of its being disposed to exhibit causal patterns that map onto inferential patterns among propositions. Might Farkas and Masrour have something similar in mind? I don't think so. CRT works as a theory of N-I properties only for a system that can enter states expressive of propositional attitudes. Thus transitions among states count as *inferential* transitions, and certain items within those states can be implicitly defined thereby. But Farkas-Masrour reductionism is not like this: while phenomenal covariance involves transitions, these transitions are between somatosensory states, not between propositional-attitude states.

Perhaps this difference highlights what is lacking in the reductive story so far: while no amount of phenomenal covarying can constitute intentional content on its own, it *can* so constitute if some sort of representational scaffolding is already in place in consciousness. Perhaps Masrour's schematic representations could be conscripted to make up the difference. Perhaps, that is, schematic representations not only select *which* patterns determine content but also signal to the subject *that* these patterns determine

content. [11] The idea here would be that phenomenal patterns constitute intentionality for a subject when the subject is otherwise prepped to take them to do so.

Now, I am not sure the resulting picture is clear. More problematically, I am not sure the resulting the picture counts as reductive, since something primitively representational (if only very schematically so) has been added to it. But even setting these worries aside, a second explanatory question needs answering: (2) what makes it the case that phenomenal covariance of a certain sort constitute a particular intentional content and not another? Suppose that two super-cognizers were able to keep a running list of all the S-S properties instantiated by a human subject over a sufficiently long time-interval. Suppose further that the super-cognizers were informed (a) that some of the patterns of instantiation fix intentional content; and (b) which patterns in particular do so. And then suppose they were asked to *interpret* the current phenomenal state of the subject. How divergent of interpretations might the two super-cognizers come up with? I want to say: *massively* divergent. While one might treat the subject as representating persisting objects in her environment, the other might treat the subject as representing *momentary* objects in her environment, or *past* objects in her environment, or gappy objects scattered throughout space and time, or temperature changes on the surface of the sun. And if nothing in the picture could rule out such bizarre interpretations as inapt, nothing in the picture could rule them out as *determinately not what is presented to the subject*. The only way I can think to rule them out would be to again expand the job-description of schematic representations. Perhaps schematic representations could supply

---

[11] It is interesting that Masrour does not call the primitive phenomenal elements "sensations" but rather "representations of some properties." (See the block quote above.) If he is using a functionalist notion of representation here, fine; but I worry that he is assuming that a subject *takes* the primitive phenomenal elements *to be* representations. But the question at issue is when/why subjects take phenomenal elements to be representational.

the subject with a means of interpreting the phenomenal patterns she is presented with.

But if that is the picture, then phenomenal patterns aren't constituting content; rather, schematic representations are grounding content by supplying an interpretation of the patterns. And now a similar dilemma besets the metaphysical rendering of reductionism to the one that besat the epistemological rendering: either schematic representations are conscious or they are not. If not, the reductionist has dispensed with the Perceptual Grounding Thesis. If yes, then the reductionist has appealed to an unreduced instance of conscious intentionality, and her theory is incomplete.

In sum: Farkas and Masrour propose that phenomenally-determined intentionality can be reduced to phenomenal covariance, but they have yet to fully carry out the reduction. First, there is no clear way to extend the proposal to the imaginative domain. Second, their explanatory story (in terms of *constitution*) is obscure even in the domain of the perceptual, where phenomenal covariance makes the most sense. Now, even if the reductionist strategy pursued by Farkas and Masrour fails, it would not follow that there is no way to reduce phenomenally-grounded intentionality to a set of S-S ingredients. But I don't think there are any other resources at a would-be reductionist's disposal, besides those we have already deployed in the effort to make sense of the Farkas-Masrour account. First of all, it seems that any attempt at reductionism will be pushed to invoke something like phenomenal covariance, since anything simpler is obviously inadequate. (A mere *accumulation* of S-S properties is never going to deliver anything intentional— that's for sure.) Second, it is hard to imagine potentially relevant types of or additions to phenomenal covariance that we have not already considered. Specifically, we have discussed phenomenal patterns at a time and over time; subsets thereof; anticipations of

successive iterations thereof; non-conscious interpretations thereof. None of these refinements have helped. The reductionist appears to be out of options. Hence, if phenomenal states ground intentional states, then some phenomenal properties are primitively intentional. P-I properties amount to a *sui generis* type of phenomenal property.

3. The Mechanics of Phenomenal Grounding.

Going primitivist (rather than reductionist) about P-I properties engenders problems of its own. For one thing, it might be thought that primitivism implies a crazy explosion of primitive phenomenal properties: we can consciously think about any individual or category we can demonstrate or assign a name to; yet it is implausible that there is a primitive, proprietary phenomenal property corresponding to every such intentional content. Call this "The Abundance Problem."[12] For another, if we could be *acquainted* with the content of every intentional state of ours, then we should expect the same quality of grasp of our thought-contents as we have of other phenomenal properties such as pain. But the difficulty (some would say *failure*) of conceptual analysis shows that this is not the case: our grasp of our thought-contents is quite poor. Call this "The Analysis Problem."[13]

To explain how P-I properties partly ground all intentional properties in a way that avoids these problems, I will use as a jumping-off point the semantic theory developed by Bertrand Russell in "Knowledge by Acquaintance and Knowledge by Description." For Russell, propositional attitudes are polyadic relations between a subject

---

[12] This worry was suggested to me by Masrour in conversation.
[13] See Pitt (2011b).

and the various constituents of a proposition. If, for example, the relation *judging* holds

between a subject, Ann, love, and Bill (in that order), then the subject judges that Ann

loves Bill.[14] At the same time, Russell holds that acquaintance is the source of semantic

understanding: "Every proposition which we can understand must be composed wholly

of constituents with which we are acquainted."[15] Russellian acquaintance is a relation that

holds between a subject and an object of one of two types: a mental particular (he calls

them "sense-data") or a universal. But this raises a puzzle: surely we can make judgments

about objects with which we are unacquainted (in Russell's sense): unless Ann and Bill

are mental particulars, it would seem that the subject cannot make judgments about them,

since she cannot be acquainted with them. And this isn't just an issue for particulars;

Russell acknowledges further that there are universals with which we are unacquainted,

but about which we can make judgments.

    Russell's solution to this puzzle is to introduce *descriptions* into the picture. In

order to make a judgment about an object with which we are unacquainted (either a

particular or a universal), a subject needs to be able to construct a definite description of

that object out of objects with which the subject *is* acquainted. To use Russell's example,

if an associate of Bismarck wished to make a judgment about him, he might construct a

description out of certain sense-data "which he connected (rightly, we will suppose) with

Bismarck's body." We who have never met Bismarck will have to employ objects further

removed from him: "When we, who did not know Bismarck, make a judgment about

him, the description in our minds will probably be some more or less vague mass of

---

[14] Russell gives his example using the variables 'A' and 'B'.
[15] Russell (1910), p. 117.

historical knowledge—far more, in most cases, than is required to identify him."[16] Thus the propositional thoughts each of us has about Bismarck will typically be much more complicated, and idiosyncratic, than the sentences we would use to communicate those thoughts.

I reject a number of elements of Russell's picture. For one thing, I do not believe in Russellian mental particulars; for another, I do not believe that we bear Russellian acquaintance to universals.[17] I endorse Russell's core semantic intuition, however: all intentional contents are either directly grasped by a subject or else indirectly picked out via something like a description constructed from directly grasped elements. With P-I properties in the picture, we have a way to endorse Russellian semantics without endorsing Russell's metaphysics. P-I properties present an intentional content to the subject. They can thus explain how it is that a subject "directly grasps" an intentional content. And if they are able to semantically combine with one another in the right sorts of ways, they can form structures expressive of descriptions.

Turning from semantics to metaphysics: we are now in a position to understand how phenomenology grounds intentionality. I take the relevant grounding-relation to be *realization*. Take some intentional property, *being about F*. This property is multiply-realizable. One way for it to be realized is for a subject to instantiate that P-I property whose intentional object is F; the subject's phenomenal state thereby realizes the intentional property *being about F*. A second way for this intentional property to be

---

[16] Ibid., p. 114-15.
[17] I do believe we can bear a type of acquaintance to universals. When we introspectively attend to phenomenal states that include P-I properties, we can become acquainted with the intentional objects of those P-I properties. This is not Russellian acquaintance because it is not an unmediated relation between a subject and an extra-mental entity.

realized is for the subject to instantiate a structured complex of P-I properties that expresses a description that picks out F.[18]

If the description so expressed contains an indexical or demonstrative (e.g. "that thing there with such-and-such features"), the description will conscript into semantic service whatever extra-mental reality answers to the demonstrative or indexical. In such cases, the realizer for the intentional property *being about F* will include bits of extra-mental reality. Indeed, there are as many potential ways for *being about F* to be realized as there are ways for the subject's mental state to be about F. It may be the case that most of the intentional properties I instantiate are realized in part by extra-mental goings-on, and some may be realized by an enormous amount of extra-mental goings on. (Example: "Let's fill our glasses with *whatever filled the glasses of the kings and queens of that country Julia visited last year*.") But they will always be partly realized by the P-I properties I instantiate. The P-I properties I instantiate combine to form the *modes of presentation* of all the intentional contents I can entertain. (That is the basic framework, at any rate. I will have much more to say about the details in chapters 4 and 5.)

Thus the primitivism I advocate regarding P-I properties is a *sparse* primitivism. And sparse primitivism is able to avoid the Abundance Problem and the Analysis Problem introduced at the beginning of this section. As regards the Abundance Problem: There is no proprietary P-I property for every possible thought-constituent; rather, P-I

---

[18] In chapter 1 I raised and then rejected the possibility that N-I properties could be understood as *realizers* of intentional properties. I suggested instead that they should be understood as *species* of intentional properties. If N-I properties cannot be realizers of intentional properties, how is it that P-I properties can? Why not say that P-I properties are likewise *species* of intentional property?

My answer to the first question is that, in the case of P-I properties, it is explanatorily transparent how it is that P-I properties realize intentional properties, whereas no such explanatory transparence is forthcoming for N-I properties. My answer to the second question is that I deny that N-I properties are species of intentional properties. They are species of functional-dispositional properties, and that is all that they are.

properties from a sparse set, deployable in countless combinations and contexts, can pick out countless kinds and categories. As regards the Analysis Problem: it is no surprise that conceptual analysis is elusive, so long as intentional contents can be fixed by indexical- or demonstrative-involving modes of presentation, and hence fixed in part by the extra-mental goings on conscripted by those modes of presentation.[19]

Let us suppose that the idea of a P-I property is coherent and that the way in which P-I properties ground intentional properties is clear. One residual issue remains: why think that *all* intentional properties are partly grounded in phenomenal properties? Couldn't there be non-phenomenal realizations of intentional properties? There is, after all, nothing in the *concept* of an intentional property that guarantees that all instantiations of such properties are phenomenally grounded: the concept of intentional directedness and the concept of phenomenality are not obvious intertwined.

My answer is that I take the inwardness intuition of Searle and McGinn (about the essentially subjective nature of intentional phenomena, discussed in the previous chapter) to be a good one. I mentioned three interpretations of the intuition: (1) Intentional properties are not essentially inwardly-directed but acquire the property of being inward-directed when brought to consciousness. (2) Intentional properties are not essentially inwardly-directed but their instantiation-conditions guarantee that they are necessarily inwardly-directed. (3) Intentional properties are essentially outward-directed and inward-directed. I advocate the second interpretation. That all intentional properties are subjectively presented is a metaphysically necessary truth, analogous to the truth that *anything that has a color has a shape*. There may be some phenomenon worthy of the

---

[19] Another obstacle to conceptual analysis is the fact that many of our concepts are prototype-based. I develop a phenomenalistic theory of prototype-based categorization in chapter 5.

name 'aboutness' or 'directedness' according which that is not so, but it would not be *semantic* aboutness, the sort of aboutness or directedness exploited by thinking things. Now, to say that all intentional phenomena are essentially subjective is not to say that all intentional contents are *narrow*. Intentionality is not *exclusively* a subjective phenomenon but it is *necessarily* a subjective phenomenon.

4. PGT and Nonconscious Intentionality: Common Sense

PGT, it might be objected, suffers from an obvious limitation: it ignores the phenomenon of non-conscious intentionality. If all intentional properties are partly grounded in phenomenal properties, then it is not possible for an intentional property to be instantiated without bearing some very close connection to the instantiation of a phenomenal property. Yet it is apparent that cognitive systems do instantiate intentional properties that have no such connection with occurrent phenomenality. There are at least two reasons to posit such states. First, *common sense* tells us that belief-states (and perhaps some other standing states) persist through changes of phenomenology. For example, it seems obvious that a theist does not cease to believe that God exists while her attention is otherwise than theologically engaged. Second, *cognitive science* tells us that informational processes underlie our conscious lives. Vision-processing, for example, involves performing computations on visual representations, the output of which are conscious states but whose intermediate steps are never brought to consciousness. I will deal with the first topic in the present section and the second topic in the following section.

I follow Searle in holding that attributions of standing states such as beliefs (to which I'll restrict my discussion in what follows) can be literally true, and that the truthmakers for such attributions have something to do with *a subject's dispositions to make conscious judgments*. (My proposal is importantly different from his in a few details, as we will see.) Searle calls this component of his view "the Connection Principle," which he states as follows: "All unconscious intentional states are in principle accessible to consciousness." This way of putting things has puzzled some readers.[20] In brief, such puzzlement stems from Searle's explicit statement that non-conscious states are physiological states of the brain; consequently it is hard to know what it is for such states to be "accessible to consciousness," since physiological states of the brain, as such, are never accessible to consciousness, in the sense of being *introspectively* accessible (and no other sense of 'accessible' comes to mind).

Such puzzlement can easily be cleared away, however. Elsewhere Searle makes it clear that he is using "unconscious" and "non-conscious" differently: while "non-conscious states" are occurrent states of the brain, "unconscious states" are *dispositional* states of the brain, i.e. dispositions to generate conscious states. Searle is not thinking of unconscious mental states as token- or even type-identical to conscious states. Rather, "The ontology of the unconscious consists in objective features of the brain capable of causing subjective conscious thoughts."[21] For Searle, that is, S's brain-state counts as S's belief that p if it is disposed to generate a conscious judgment in S that p.

I do not think the view will work, as it stands. There are two problems, one pertaining to the *bearers* of the relevant dispositions and the second related to the

---

[20] See, e.g., Fodor & Lepore (1994).

[21] This clarificatory claim has been neglected by readers of Searle, despite its being found only one page after his introduction of the Connection Principle in Searle (1992).

triggering-conditions of the relevant dispositions. Searle says that beliefs are *brain-states* that are disposed to generate conscious thoughts. But identifying beliefs with such brain-states is to cast the belief-net both too narrowly and too broadly. Too narrowly, because it seems unduly parochial to think that only brainy systems can have beliefs; perhaps there are "exotic", non-brainy conscious systems. (After all, the driving thought is that whatever systems are disposed to have conscious judgments are systems that can have beliefs; it is certainly not a priori apparent, and perhaps not true, that only brainy things are so disposed.) Too broadly, because it is conceivable that some state of my brain could be re-wired so as to generate a conscious judgment in *you*, and *not* to do so in me. So we need a different way to think about the bearers of the dispositions that serve as truthmakers for belief-attributions.

The trouble is that once we abandon brains as the bearers, it becomes difficult to find an alternative that does not multiply beliefs absurdly. Lots of things are causally relevant to my having conscious thoughts, after all: the items I look at, the books I read, the people I trust. But even if these items are disposed to generate conscious judgments in me, they are not themselves beliefs—mine or anyone else's.[22]

As a solution to this problem, Kirk Ludwig (1996) proposes, first, that we identify beliefs not with the *bearers* of dispositions but with the dispositions themselves; second, that we add the constraint that the bearers of the dispositions must be properly related to the conscious episodes they generate. In particular, the only dispositions that count as my beliefs are the dispositions had by entities necessary for the existence of the conscious judgments they generate in me. For example, while a wall may produce a perceptual

---

[22] I note that advocates of the Extended Mind Hypothesis might want to *ponens* where I *tollens* here. At the very least, I gather that something very much like the problem I am addressing is a common motivation for the hypothesis. See Chalmers & Clark (1998). For helpful discussion, see Gertler (2007).

judgment in me (by way of my having a perceptual experience of that wall), the wall does not form part of the supervenience base of my having a perceptual judgment about it.[23]

Ludwig's proposal goes a goodly distance in distinguishing between those dispositions that count as *my beliefs* and those dispositions that merely causally contribute to my making judgments. His view may be too restrictive, however: it rules out the possibility of a system whose states fit the following profile: (1) they are causally well-integrated within a cognitive system; (2) the conscious judgments they generate are part of the single stream of consciousness generated by that cognitive system; but (3) they are not part of the supervenience base of the conscious judgments they generate (i.e. such judgments are causally but not ontologically dependent on them). I do not see a reason to deny that such states are possible, so I do not think Ludwig's account is adequate.[24]

I doubt that we are going to find an account of the *bearers* of the relevant dispositions that is neither overly broad nor overly narrow. Consequently, we will need a different way to characterize non-occurrent intentional states. I propose that we identify them not with the disposition to *generate* conscious judgments but rather with the disposition to *entertain* conscious judgments. The bearer of such a disposition is the selfsame subject who consciously judges. Note that we can beg off any further commitments about what a subject is, in our account of what it is to have a belief.

The second problem for Searle's strategy has to do with the *triggering-conditions* for the relevant dispositions. There are many conscious judgments that I am disposed to make, so long as I am stimulated in certain ways. I am disposed to judge that Obama is the U.S. president if you ask me to name the current occupant of that office. I am also

---

[23] Ludwig (1996), footnote 24.
[24] My view of ontological emergence vindicates such a possibility. See chapter 6, section 2.

disposed to judge that Obama is in China, if you show me a recent newspaper that says as much—yet I do not currently believe that Obama is in China. I am disposed to judge that Obama is younger than my furnace repairman; while I have never before entertained the proposition and might require some reflection on my part in order to evaluate, it presumably follows pretty obviously from other things I *have* entertained and consequently counts as a belief of mine. Then again, I may be disposed to judge that Obama is a material thing distinct from but spatially coincident with the body that composes him, were I to be shown that others beliefs of mine (about which I feel quite confident) a priori entail this view. It could turn out, that is, that I am *committed to it* given other things I believe, but I would *not* count it among my beliefs at present.

Searle's account of beliefs generates puzzles along a few other dimensions as well (not having to do with the triggering-conditions exactly). Consider: I might be disposed, *on Mondays and Wednesdays*, to judge that Obama is a material thing distinct from but spatially coincident with the body that composes him; while being disposed, on *Tuesdays*, to judge that he is an immaterial soul; on *Thursdays*, that he is a worm in spacetime; on Fridays through Sundays, to refrain from judgment. It would be inaccurate, I gather, to unqualifiedly attribute to me any belief in particular regarding philosophical anthropology; a requisite measure of *stability* in my dispositions to judge is lacking. Again, consider: I may be disposed to judge that Obama is a material thing distinct from but spatially coincident with his body, but with a much lower feeling of certitude than that which accompanies my judgment that he is president—perhaps with a feeling of certitude that corresponds to a credence level barely over 50%. It would be at best

misleading to attribute to me the corresponding belief; a requisite measure of *confidence* in the judgments I am disposed to make is lacking.

What these cases demonstrate is that belief-possession is vague, where the vagueness is most likely to be explained semantically (i.e. stemming from wiggle-room, in multiple dimensions, with respect to our concept of belief). If I am *stably* disposed to *confidently* judge that p *under conditions of minimal epistemic stimulation*—e.g., I am asked to venture a verdict on p and I am supplied with no new evidence—then I believe that p; and if I am not so disposed, then I do not believe that p. But if my disposition is somewhat unstable, my judgments are somewhat unconfident, or I require a bit of epistemic pressure to induce a judgment, then it will be vague whether I believe that p. (For example, I suspect that it is vague whether Meno's slave has—prior to being questioned by Socrates—the geometrical beliefs relevant to calculating the area of a square.)

Beliefs, then—as examples of non-occurrent intentional states—can be accommodated with the framework I have been developing in this chapter: a subject believes that P iff she is stably disposed to consciously make a confident judgment that P under conditions of minimal epistemic stimulation. Now, there is a type of non-occurrent intentional state that does not admit of such treatment. These are what we might call "Freudian beliefs," i.e. intentional states that motivate a subject's behavior despite the fact that she would not be disposed to make the corresponding judgment, or may even be disposed to make the *opposite* judgment. Here is an example:[25] a professor is disposed to judge that foreign candidates are of equal competence as domestic candidates. Yet when she reviews applications for an opening in her department, she is much more quickly

---

[25] From Peacocke (1998).

dismissive of foreign candidates than she is of domestic ones. There is a sense in which she believes that foreign candidates are less capable, despite the fact that she would explicit affirm the negation of that proposition.

I do not think that it would be literally true to attribute to the professor the belief that foreign candidates are less qualified. A closely related claim *is* literally true, however, viz. the claim that the professor behaves *as if* she believed that foreign candidates are less qualified. I take "S behaves as if S believed that P" to mean that S's behavior is most straightforwardly *rationalized* in terms of a belief that P. More precisely: were it the case (contrary to fact) that the behaviors of S in question were caused by a conscious intention, the conscious intention that would explain those behaviors (a) most parsimoniously and (b) involving the least amount of deviation from S's beliefs would include the belief that P. Note that (a) and (b) are apt to recommend distinct accounts of the content of P, and choosing between them will be a matter of convenience rather than truth; consequently, justification for attributions of Freudian beliefs involves a pragmatic component.

Now, this suggestion is likely to sound ad hoc, in the absence of a general theory of the relationship between intentional states and behavior within which it can be located. There is no space to develop such a theory in full here, but I can give the sketch of one. Here is the basic idea: rational behavior is caused, in the first instance, by conscious intentional states.[26] These could be as rich as pairs of conscious judgments and conscious desires; or they could be more impoverished, such as the perceptual representation of something and a concomitant aversive feeling toward that thing. Now suppose that type-

---

[26] By 'rational behavior', I mean behavior that is caused or guided by manifestations of intentional capacities, broadly construed.

identical intentional states repeatedly cause the same behaviors. It's plausible that the result will be a kind of habituation that bypasses consciousness altogether. This might occur as follows: causal connections get built up between (a) the neural bearers of the disposition to entertain to the relevant intentional states, and (b) the relevant behavioral output. The conscious mind thus "programs" or "rewires" the brain, thereby sublimating the causal role of an intentional state to the level of the neurons. Empirical findings confirm that something very much like this occurs: voluntary behaviors, including very complicated ones, become automatic over time, and the details of how they are produced become less and less consciously accessible.[27] Now, the resulting habitual behaviors are still attributable *to the agent* (at least if she is still in a position to consciously preempt them), but they are rational only in a derivative sense: while their causal ancestry includes conscious intentional states, their immediate causes include only brain-activations (*programmed* by past intentional states).

Suppose the professor in our example has acquired a behavioral habit, via the process of sublimation just described: somewhere along the way she habituated herself to distrusting foreigners. (Such a process could have occurred at an arbitrarily early stage of life—so long as she had developed cognitively such as to able to consciously represent the category *foreigner* or something close enough extensionally). But suppose she later becomes convinced that foreigners are no less trustworthy than her fellow nationals. Though her habitual behavior remains—and though it is in some sense a real expression of foreigner-aversion—it would not be accurate to say that *she believes that foreigners are untrustworthy.* Note, further, that if brains can be programmed by conscious minds, then they can presumably be programmed by gene-expressions or by operant

---

[27] Cf. Baars & Gage (2010), p. 52.

conditioning or by quantum accident. In such cases, behaviors resulting from such

programming do not count even as *derivatively* rational, since they issue from

programming that lacks the right sort of history—yet it may be apt to account for such

behaviors in terms of Freudian states. Hence an agent's behavior is at best an indirect and

fallible guide to what she actually believes (i.e. what she is disposed to consciously

judge). It may be useful for us to deploy a different notion of belief, one that is bound up

with rational behavior and not with conscious judgment.[28] But this notion is no *more* than

useful. It need not answer to objectively real, non-conscious intentional states.

5. PGT and Nonconscious Intentionality: Cognitive Science

There is no common-sense notion of non-conscious intentionality that poses a threat to

PGT, then. Still, it is often suggested that the positing of non-conscious intentional states

is indispensible for theorizing within cognitive science; and as the posits of our best

science serve as our best guide to what types of things there are (a doctrine famously

averred by Quine in 'On What There Is'), we ought to believe in non-conscious

intentional states. Even some philosophers who are otherwise sympathetic with the

motivations for PGT discussed in the last chapter have concluded, on empirical grounds,

that PGT must be false. For example, Declan Smithies (2012) writes:

> [PGT] conflicts with the explanatory role that unconscious intentional states play
> in commonsense psychology and scientific psychology alike. For instance ...
> computational explanations in cognitive science appeal to computational
> processes defined over unconscious mental representations, such as Chomsky's
> (1965) tacit knowledge of syntax and Marr's (1982) primal, 2.5D, and 3D sketch.

---

[28] Unsurprisingly, one important use of the notion of Freudian belief is in psychotherapy. If I am right about
the way that sublimation occurs, it is possible for a subject to cognitively override the nefarious functioning
of a Freudian belief—no matter whether its source was itself rational or merely natural or accidental—with
an overt belief.

Some proponents of [PGT] claim that, strictly speaking, there is no unconscious intentionality, although it can be useful to speak metaphorically as if there were. However, unconscious intentionality seems to play an indispensable role in psychological explanation in common sense and cognitive science. Moreover, we have good reasons to believe in existence of these states, rather than regarding them as fictions, insofar as they play an indispensable role in explanation.[29]

I suspect that something like the foregoing line of thought has motivated many philosophers and cognitive scientists to treat Searle & his ilk with some contempt: not as providing conceptual clarification on behalf of cognitive scientists but as ignoring or even interfering with the practice of cognitive science.

I do not think that the case (for unconscious intentional states) from cognitive science is as straightforward as Smithies and others think it is. That is, the success of computational explanations in cognitive science does not commit us to believing in such states, because they fail to be indispensible *in the sense required* for the Quinean methodological principle to apply.

The first thing to keep in mind is that not all paradigms within cognitive science posit unconscious mental representations of the same sort and to the same degree. Certainly, explanations that draw on the resources of classical computation require that there are many such (per Jerry Fodor's slogan: "No Computational Without Representation!")[30]. It is less clear what role representation plays in connectionist computation. And many who work within dynamical systems and enactivist paradigms explicitly distance themselves from talk of representation. Andy Clark has gone so far as

---

[29] Smithies (2012), p. 354. Smithies adds that the empirical details are unimportant because the evidence suggests at least the possibility of unconscious intentionality. But that's a non-sequitor. We can have evidence for impossible scenarios. It is, e.g., an epistemic possibility that particle theories of light are incompatible with wave theories of light. They certainly *seem to be* incompatible. The fact that we have evidence for both does not itself show that they aren't.

[30] Fodor (1975).

to contend that cognitive science is "involved in an escalating retreat from the inner symbol."[31] Cognitive science is a young science, and its leading practitioners disagree regarding some of the most basic questions as to how it should be practiced; consequently the question of which theoretical resources are indispensible to cognitive science is, I gather, rather difficult to answer.

Let us suppose, though, that representation-heavy computationalism remains central to the practice of cognitive science. I take the computationalist approach to cognitive science to amount to something like the following: the endeavor to provide computational models of those processes that mediate the inputs and outputs of a cognitive system. Classical models invoke the creation, retrieval, and transformation of explicit symbols according to algorithmic rules. Connectionist models invoke units with weighted connections between them. I will first explain why I do not think classical-computational explanations pose a challenge to PGT and then I will briefly comment on the implications of my discussion for connectionism.

It has often been observed (since David Marr [1982]) that classical computational explanations admit of a tripart structure. Here is one way to think of the three "levels". At the *cognitive* level, a computational process amounts to a series of transformations over meaningful symbols governed by explicit rules. At the *syntactic level*, a computational process amounts to a series of transformations of shapes or structures in accordance with algorithms. At the *implementation* level, a computational process amounts to a series of mechanical transitions in whatever medium the implementing cognitive system is made of.

---

[31] Clark (2001), p. 121. See chapters 2, 4, 7 and 6 for helpful discussions of the respective paradigms just mentioned.

These levels provide answers to different sorts of explanatory questions. Suppose you want to know how an organism went about solving a particular cognitive problem (choosing one grapefruit from a pile of grapefruit, recognizing a friend, doing "mental math," etc.). With a classical computational model as your guide, you can look for whatever causal process in the organism connects the relevant inputs to the relevant outputs according to the functional pattern specified by the model. "Aha—the organism deployed a causal process involving structures along *that neural pathway*," you can conclude. In other words, your computational model points you toward a causal explanation at the implementation level. But you may want to ask a more general question. You may want to know, for example, what *type of process* an organism uses to solve a certain type of cognitive problem, even if (owing to neural plasticity, perhaps) the organism implements this process differently at different times. Or you may want to know what type of process organisms *of the same type* use to solve a cognitive problem. To answer such questions, you'll want to extract from your computational model a causal explanation at the syntactic level, whereupon you may conclude: "Aha—organisms of this type deploy causal processes involving structures with *that syntactic shape*."

What is important to see is that there are no explanatory questions for which *the cognitive level* supplies a causal-explanatory answer. No matter the level of generality of an explanatory question, computational models point to syntax as that which is causally salient. What the symbolic structures *mean* never comes into causal play. Consequently, semantic descriptions of the transformation-rules that "govern" the system never come into causal play, because those descriptions are apt only derivatively off the meanings of the symbols. But whatever mechanism performs symbol-transformations in a computer is

126

not causally responsive to the semantic properties of the symbols. (Analogy: a mechanical coin-sorter can be described as sorting coins according to their denominations, but of course the denominations of the coins do not play a *causal* role in the sorting.) So, semantic properties of the structures in a computational system are at every level epiphenomenal.

Now, I am not making a new observation about computationalism. Not only am I not making a new observation, I am pointing to a feature of computationalism that its advocates explicitly take as a key *virtue*. The thought here is that computationalism squares the apparently non-mechanical nature of intentional properties with a mechanical conception of mental functioning. Here, for example, is Fodor's sales pitch for computationalism (1987):

> Here, in barest outline, is how the new story is supposed to go: You connect the causal properties of a symbol with its semantic properties *via its syntax*...It's easy, that is to say, to imagine symbol tokens interacting causally *in virtue of* their syntactic structures. The syntax of a symbol might determine the causes and effects of its tokenings in much the way that the geometry of a key determines which locks it will open....the semantic relation that holds between two symbols when the proposition expressed by the one is entailed by the proposition entailed by the other can be mimicked by syntactic relations in virtue of which one of the symbols is derivable from the other.[32]

Fodor thinks, that is, that the general shape of belief/desire psychology is vindicated by computationalism, owing to the role that the category of syntax plays in computationalism: syntactic entities carry intentional properties, as it were, through the causal networks of a cognitive system. Andy Clark puts the point succinctly: "Reason-guided action, it seems, makes good scientific sense if we imagine a neural economy organized as a syntax-driven engine that tracks the shape of semantic space."[33]

---

[32] Fodor (1987) pp. 18-19.
[33] Clark (2001) p. 16.

Now, to say that intentional properties play no causal role in computational theories of cognitive is not to say that they play *no* role. There are, I gather, at least two types of generalizations, important to cognitive science, that quantify over intentional properties. First, intentional properties serve to individuate cognitive processes.[34] For example, a process is a decision-making process, rather than something else, owing to its output, which is a decision (understood either as an experiential state or a behavior). Consequently, just as cognitive scientists might want to quantify over instances of the same syntactic type that are differently implemented, so cognitive scientists might want to quantify over instances of the same cognitive type that involve different syntactic processes. (Example: "All mammalian decision-making involves at least one of three distinct processes.") Second, mental symbols do *have* representational contents, according to computationalism. Consequently a cognitive scientist might wish to quantify over mental representations, as individuated by their content. (Example: "Representations of physical space play a role in mammalian processing of vision and audition.")

We are now in a position to respond to the charge that unconscious intentional states amount to an indispensible posit of classical-computational models, and hence that we are committed, according to our best philosophical methodology, to believe in them. I say that we are *not* so committed, and here is why. Quine's methodological guide to ontology—*believe in the posits of our best science*—need not be treated as our *only* guide. Other guides include Occam's Razor and the Eleatic Principle (to be is to be a potential cause). Taken together, these principles recommend the following emendation of Quine: all other things being equal, we ought to be ontologically committed to those entities, posited by our theories, that play an indispensible *causal* role in our best

---

[34] For helpful discussion of this point see Shagrir (2001).

scientific theories. But unconscious intentional states do not play an indispensable causal role in classical-computational explanations. They may be indispensible for certain purposes, viz., for stating certain generalizations. But this is a merely pragmatic indispensability, i.e. the only way to succinctly describe certain patterns.

It may be objected that computationalism construes the semantic properties of mental representations as causally efficacious after all. Consider: what exactly is it for a mental symbol to have a representational content? A cognitive scientist could be silent on the matter, of course. But if pushed, she could either hold that mental symbols' representational contents amount to irreducibly intentional properties of those symbols, or that they amount to the N-I properties of Chapter 1 of one flavor or another. I take the latter route to be more in keeping with the spirit of computationalism. N-I properties are dispositions of their bearers to be activated by the system in certain ways under certain conditions, and dispositional properties are presumably *not* causally epiphenomenal.[35]

But there are two distinct reasons that the introduction of N-I properties as a way of cashing out representational content poses no new challenge to PGT. First, the effects that these properties dispose their bearers to bring about cut across the causal relations relevant to computational explanation. Essentially, they dispose their bearers to participate in certain computational processes. But these dispositions are not themselves causally relevant to the computational processes in which the disposition-bearing mental symbols appear. It might be argued that N-I properties, though not indispensible to classical computationalism in the sense of playing a causal role in classical-computational explanations, are nonetheless indispensible in yet a third sense of 'indispensibility' (besides causal indispensability and pragmatic indispensability): they

---

[35] Though the matter has been disputed. See Prior, Pargetter, & Jackson (1982) and Schaffer (2003).

are needed as a way of a squaring the ontology of cognitive science with the ontology of the physical sciences. But recall that (and here is the second reason), per the Acquaintance Argument of chapter 1, N-I properties are not paradigmatic intentional properties. N-I properties and P-I properties are not species of the same genus. I can grant that properties of the sort specified by N-I theories are instantiated. But that fact is irrelevant to the credentials of PGT.

A brief word about connectionism, as promised. Connectionist architectures differ from classical architectures in that they employ parallel processing and do not involve discreet symbols. But the causal status of cognitive-level descriptions of connectionist models is the same as that of classical models; syntax (of a different sort) does all the causal work. Connectionists may very well quantify over a connectionist network's representational contents for purposes of making certain abstract generalizations. But the mere pragmatic indispensability of talk of unconscious intentional states does not call for ontological commitment to such states.

I conclude that the general success of computational models in cognitive science does not itself give us good reason to believe in non-conscious intentional states. PGT may yet be subject to pressure from empirical science, however, by way of *specific empirical findings*. I'll discuss three findings that are readily and naturally interpreted as indicating the presence of non-conscious intentional states.

1. *Blindsight*. Some patients with damage to primary visual cortex report having no visual awareness. Yet when they are presented with a stimulus—an X vs. an O, or a horizontal vs. a vertical line, they are able to guess, with a high degree of reliability,

which manner of stimulus has been presented.[36] A natural explanation is that visual information has been recorded and made available for verbal report, despite never reaching consciousness.

2. *Subliminal semantic response*. Subjects respond differently to coherent /familiar vs. nonsensical/bizarre semantic cues, even when those cues are not part of conscious visual awareness. One recent study illustrates this phenomenon vividly.[37] Visual experience is the result of the combination, in visual cortex, of the activation of processing streams from each retina. Rarely do we notice the unique trace of each independent input (though we can sometimes attend to the "doubling" of non-focal objects[38]). But when the input streams are sufficiently divergent, conscious experience will selectively attend to one stream rather than the other, a phenomenon typically referred to as 'binocular rivalry.' It is possible to rapidly stimulate one retina while under-stimulating the other and hence to build into the design of an experiment which retinal input will reach visual awareness and which will not—an experimental paradigm called "Continuous Flash Suppression," or "CFS." Sklar et al used CFS in following way: while stimulating a subject's retina using CFS, they presented to the other retina a written phrase, first very dimly and then increasingly vividly. They asked their subjects to indicate the moment that they became visually aware of the phrase. The results were interesting: subjects became aware of nonsensical sentences more rapidly than meaningful ones (e.g. 'I ironed coffee' vs. 'I made coffee' and 'the window got mad at her' vs. 'the gentleman got mad at her') and negatively-valenced phrases more rapidly than neutral ones (e.g. 'black eye' vs. 'sand box'). A natural explanation is that the

---

[36] Weizkrantz et al (1974).
[37] Sklar, et al (2012).
[38] For interesting discussion see Schwitzgebel (2011), pp. 30ff.

semantic cues were understood unconsciously, and only subsequently presented to consciousness, more or less rapidly depending on the perceived interest or importance of their content.

   3. *Split-brain disassociation*. The corpus collosom connects the two cortical hemispheres in a healthy human brain. Some epileptic patients have had their corpus collosom severed, so as to disrupt the electrical storm that overwhelms the brain as it undergoes an epileptic seizure. While the two cortical hemispheres of so-called 'split-brain' patients still interact via sub-cortical connections, their hemispheres process inputs and outputs without the constant inter-hemispheric interaction characteristic of an intact brain. This means that inputs to the right side of the visual field and inputs to the left side of the visual field seem not to be integrated in the normal way, and that behavioral outputs to the right and left sides of the body seem not to be integrated in the normal way.[39] Two split-brain studies are especially relevant to the topic at hand. In one study[40], a patient was shown a picture of a bird's foot in his right visual field and a snowy scene in his left visual field. Each hand was offered several images, printed on cards, from which to choose a suitably matching image; the right hand selected an image of a chicken and the left hand selected an image of a snow-shovel. When the patient was asked to explain the selection of the shovel, he seemed unaware of the snowy scene but instead explained that the shovel could be used to clean out a chicken-coup. Note that the left hemisphere—i.e. the hemisphere involved in connecting the bird-foot image to the chicken image—contains the patient's dominant language-processing center (as is

---

[39] Behaviors triggered by the right- and left-hemispheres in split-brain patients sometimes conflict; patients have been known to try to restrain one hand with other. See Brogaard 2012.

[40] Gazzaniga 2002. ("The Split Brain Revisited," Scientific American.)

common for humans). In a similar study,[41] when the patient was presented with the word 'hammer' in the right visual field, he could see and understand the word. When presented with the word 'saw' in the left visual field, he reported seeing nothing but (to his own mystification) was able to draw a picture of a saw with his left hand. A natural explanation of these findings is that the right hemisphere can understand visual information, including written language, but that these processes remain disconnected from conscious thought or the production of speech.

Now, to say that such explanations are natural is not to say that they are forced; as suggested in the previous section, they could be natural because convenient, allowing us make generalizations that abstract away from the mechanical details. Further, if I am right about PGT, then such explanations involve something in the neighborhood of a category error (however convenient an error), since there just are not any instantiations of genuine intentional properties save for those subjectively presented (or disposed to be). So, the question we need to ask with respect to the three empirical phenomena just described is whether they can be plausibly accounted for either by (a) invoking conscious intentionality after all, or (b) invoking complex mechanisms, at no stage of which do intentional properties play an ineliminable causal role. I think they can be so accounted for, though of course the matter is open to further empirical investigation. I will briefly discuss strategies for reconciling each of the three phenomena with PGT.

1. There are at least three competing theories about the phenomenology of blindsight. The first is that blindsight lacks any associated phenomenology: conscious states plays no role in mediating the blindseer's visual inputs and her verbal outputs, save for her ability to hear and understand a researcher's instruction to venture guesses. This

---

[41] See http://www.youtube.com/watch?v=ZMLzP1VCANo&feature=youtu.be.

account makes sense of patients' surprise at their own accuracy (though it leaves out some patients' reports of feelings of "smoothness" or "jaggedness", accompanying visual presentations of Os vs Xs). According to a second interpretation, blindseers do undergo visual phenomenology, but a phenomenology degraded to near-threshold levels such that patients struggle to attend to it and consequently hesitate to report it. Recently Berit Brogaard has ventured a third interpretation, one that I find plausible[42]: blindseers experience no *visual* phenomenology but they do experience *cognitive* phenomenology. That is, their vision system gives rise, at least under circumstances similar to those studied, to conscious judgments with a somewhat obscure content and with a moderate degree of felt certitude. Their accurate verbal responses are consciously caused after all, then. (I discuss cognitive phenomenology in chapter 5, section 6.)

2. In the CFS study, the key difference between subliminal semantic stimuli that were brought to consciousness rapidly rather than slowly was that the former were in some sense perceived to be note-worthy, important, striking. Now, it is a common phenomenon that certain sensory stimuli tend to phenomenologically "pop out," grabbing the subjects attention without her having a say in the matter, as it were. This is true with respect to items whose color and shape contrast starkly with surrounding items, with respect to the sound of one's own name, and with respect to certain other noises (e.g. the cry of a baby or the barking of a dog). What is interesting about the CRS study is that the difference between the phrases that "pop out" and those that did not is, intuitively, a matter of what they *meant*. That is why it is natural to posit non-conscious semantic understanding in order to explain the data. But so long as the visual system *mimics* such semantic understanding as "I *iron* coffee? Huh? That's odd—let's take a closer look",

---

[42] Brogaard (2011).

there is no requirement that the system actually so understand. How could the visual

system come to mimic such understanding? A natural answer is to extend the notion of

sublimation introduced above to lexical items and their semantic relationships. Suppose

that lexical items are in some manner 'stored' neutrally. (A fanciful version of this

picture would have it that a single neuron plays the role of (a) "recognizing" a word in

perception, and/or (b) "converting" a conscious representation of a word into a motor

output in the form of speech or writing.) Next, allow that consciously-grasped semantic

relationships between words (such as determinate/determinable) can likewise be

sublimated as associative connections between such neural structures. (Such associational

connections are presumably involved in how *part of speech* is neurologically registered.)

Third, allow that entire semantic "frames" can be sublimated.[43] Allow, that is, that among

the associative connections are those that map onto conceptual associations: associations

that connect lexical items in a category (e.g. the verb 'iron') with lexical items in closely

related categories (e.g. terms that canonically pick out either performers or recipients of

ironing). If there be such neural structures, then subliminal semantic stimuli can have the

effects the CFS study shows them to have, without non-conscious intentional states'

playing any causal role in the process.[44] Add that *affective valence* can be part of the

lexical associative net and we've accommodated all of Sklar et al's findings.[45]

---

[43] Semantic frames are networks of terms, where understanding the meaning of any one of the terms requires an understanding of how a chunk of reality functions. An example is the commercial event frame, whose elements include *buy(er)*, *sell(er)*, *pay*, *spend*, *cost*, *charge*. See Fillmore (1982).
[44] Sklar et al acknowledge something like this (alternative) explanation for some of their data.
[45] Sklar et al use the following reasoning, I gather: 'black eye' has its negative valence—and hence its affective urgency—in virtue of nothing short of the semantic features of the whole phrase. That is, you have to know what black eyes are, and not just what eyes and black things are, in order to know that it refers to a wound. I suspect, though, that such phrases are syntactically processed *as lexically primitive*. In other words, the fact that the stimulus is a two-word phrase does not add anything to their case for nonconscious semantic processing.

135

3. I return, finally, to the strange case of split-brain disassociation. The phenonema in question seem best described as follows: a patient engages in behavior (selecting an image, drawing a picture) that for all the world looks to be caused by relative high-level cognizing—noticing that a tool "fits" a certain scenario; knowing what an object named by a word looks like—and yet the patient reports having no awareness of such cognizing. I can think of three explanations, consistent with PGT, all of which I take to be live options. First, it could be that the lack of integration of the patient's two cortical hemispheres leads to truly disintegrated mind, such that we have a case of two subjects occupying a single body, only one of whom has the ability to speak. There are certainly data that lend plausibility to this suggestion: for example, questions directed to one or the other hemisphere of a split-brain patient have been known to prompt very different answers, answers that seem to express divergent personalities.[46] On this view, both hemispheres give rise to conscious, intentional states that are explanatorily relevant to ensuing behavior.

Second, it could be that the patient has but one stream of consciousness, no phenomenology accompanies her apparently cognitive behaviors, and that they really are only *apparently* so: the sublimation of semantic relationships in the form of complex associative mechanisms introduced in the previous few paragraphs are sufficient to explain the ability to pair two images that "fit" well and the ability to draw a picture of the type of thing named by a word.

My preferred explanation sits in between these two extremes. Split-brain patients, I propose, suffer from a peculiar form of *neglect*. Neglect involves a propensity to ignore some portion of the visual field. (Neglect-sufferers have been known to only eat half of

---

[46] See Brogaard (2012).

the food on their plate.) Often their omissions can be overcome if pointed out to them. Consequently, the phenomenon is usually understood as an *attentional* deficit (in contrast to forms of blindness or agnosia, which amount visual and/or cognitive deficits). I suggest that split-brain patients suffer from *introspective neglect*, with respect to those aspects of their phenomenal manifold generated by their right hemisphere: when prompted to report such aspects, they cannot owing to an inability to properly attend. (It is as though the relevant perceptual episodes are instantly forgotten as soon as they are experienced.[47]) Nevertheless, they do instantiate P-I states as of the relevant stimuli, which states are explanatorily relevant to their behavior—despite the fact that split-brain patients cannot attend to these P-I states, and so cannot form introspective judgments about them.

I have discussed only three alleged cases of empirically-discovered unconscious intentionality. There are other cases I could have chosen, and there will be plenty more. I suspect, though, that the strategies I have employed for finessing the present challenge to PGT will apply to many other cases as well.

6. Conclusion.

According to PGT, all instances of intentional properties are partly grounded in instances of phenomenal properties. I have offered the following proposal by way of explanation of PGT. P-I properties are sui generis phenomenal properties whose nature consists in the presentation to the subject of an intentional object. P-I properties from a sparse set combine, in perception, imagination and cognition, to form modes of presentation of all

---

[47] A vivid analogy can be fund in the sixth season of the BBC's 'Doctor Who', in which there are creatures capable of instantly removing all memorial trace of having been perceived.

the intentional contents we can entertain. This does not mean that all intentional states are occurrent phenomenal states. But standing-states such as beliefs need to be understood in terms of dispositions to generate phenomenal-intentional states. Nor is my proposal at odds with the deliverances or practices of cognitive science. The *neural bearers* of dispositions to instantiate phenomenal properties can stand in causal relations to one another and to behavioral outputs such that neural goings-on closely mimic semantic goings-on. The usefulness of describing such phenomena in terms of non-conscious intentionality need not commit us to ontologically to there being non-conscious intentionality: by the admission of cognitive scientists themselves, the semantic properties thereby attributed to non-conscious processes are *causally inert*.

So there are no immediate theoretical obstacles to our taking my theory of the phenomenal grounding of intentionality seriously. Challenges aplenty crop up when we try to work out of the details of theory, however. To that task I turn in the next two chapters.

4. The Semantic Structure of Phenomenal-Intentional States

1. Introduction.

I have been contending that the Phenomenal Grounding Thesis is true: all instances of intentional properties are partly grounded in instances of phenomenal properties. In the last chapter I began to develop a theory of *how* phenomenal properties ground intentional properties. The central idea is that phenomenal-intentional properties (or P-I properties), which consist essentially in the presentation to the subject of an intentional object, comprise a sui generis mode of phenomenal property. P-I properties ground intentional properties by fixing the mode of presentation of the intentional object whereby intentional properties are individuated.

So saying hardly clears up all questions about the nature of intentional mental states. One reason is that intentional states can be semantically complex in a variety of ways. Perceptual states, for example, do not merely present features of the world; they presents those features as being had by things. The task of the present chapter is to explore what consciousness has to be like in order to ground intentional states with the types of semantic structure we know thoughts and percepts to have.

I begin (in section 2) by discussing Jackson's problem for adverbialism, and the unity-of-the-proposition problem for theories that treat propositions as structured. Correlates of these two problems indicate the need for account of how one intentional item can be attributed to another in perception and thought, and I propose (in section 3) a mechanism that explains this form of semantic structure. I then turn my attention to a variety of other types of semantic structure that crop up in phenomenal-intentional states:

perceptual recognition and structure that spans perceptual modes (in section 4); the structure of relational contents (in section 5); and logical structure, including quantification, truth-functional complexity, and modal operators (in section 6).

If the proposals of this chapter are successful, I will have explained how the metaphysical structure of our conscious states grounds the semantic structure of the contents presented in such states. Further, I will have done so while preserving phenomenological plausibility. That is, while I am not engaging merely in descriptive phenomenology—my proposals amount to theoretical posits, rather than introspective givens—I contend that my proposals are not inconsistent with the manifest structure and occupants of consciousness.

2. Jackson's Problem and the Unity-of-the-Proposition Problem.

Suppose that all intentional properties are grounded in phenomenal properties. Suppose further that I am right about the *mechanism* of such grounding: viz., that phenomenal-intentional properties form the mode of presentation for all intentional contents. And finally, suppose that phenomenal properties are monadic properties instantiated by subjects. (That is, phenomenal properties are not relations that hold between a subject and something else.) It follows that all intentionality bottoms out in monadic properties of subjects. Call this "the monadic theory of intentionality," or *monadicism.*

There is something counterintuitive about this result. Whenever we attribute intentional states to subjects, we do so using decidedly relational language. If a subject's perceptual state has *the snake in the grass* as its intentional content, we say that the subject perceives the snake in the grass. If it is a cognitive state we're attributing, we

might say that the subject thinks about the snake in the grass. If it is an imaginative state, we say that the subject *imagines* the snake in the grass. In all three cases we use a transitive verb or verb phrase. And generally, transitive verbs are used to predicate relations: when we say that a subject *hit* his sister, *formed* a government, or *antagonized* a police officer, we are describing relations of which the subject is one relatum. We might think, then, that a *relational* theory of intentionality is the far more natural one, at least when it comes to capturing the way we talk.

Nevertheless, there are well-known reasons to suspect that grammar misleads us when we attribute intentional states. If I tell you that I have been having a lot of thoughts about G.E. Moore lately, this need not mean that my thoughts stand in the aboutness relation to a man long dead; nor does Peter Pan need to have ever existed for me to imagine him. Famously, Roderick Chisholm (1957) developed a way to translate attributions of intentional states into constructions that do not make use of transitive verbs. He recommends treating transitive mental-action verbs as disguised *intransitives*, in connection with adverbial modifiers rather than with direct objects. Using 'sense' as the most generic mental-action verb (as roughly equivalent to 'experience'), Chisholm suggests that 'S senses an F' can be translated as 'S senses F-ly'. The result, while linguistically unnatural, is more ontologically perspicuous than our ordinary way of talking. This approach has come to be known as 'adverbialism.'[1]

Frank Jackson (1975) has pressed an important objection against adverbialism. His concern is that the adverbial analysis lacks the grammatical tools to capture an obvious characteristic of mental states, viz., that the features they present can be

---

[1] 'Adverbialism' is often used as a name for what I am calling 'monadicism'. (See Kriegel 2011, for example.) I resist such usage. Intentional states can be monadic but not adverbial in any obvious sense; intentional state *ascriptions* can be adverbial but not monadic in any obvious sense.

structured in certain ways and not others. For example, suppose a subject's perceptual

state has as its intentional content a red square and a blue circle. There seems to be no

way to translate "S senses a red square and a blue circle" using Chisholm's adverbialist

schema such that the *structure* of the subject's perceptual state is captured by the

grammar of the resulting translation. Here are three options, which Jackson takes as

exhaustive: First, we could simply *conjoin* the adverbial modifiers in a list: "S senses

redly and squarely and bluely and circularly." But this can't be the right translation,

because it fails specify which property goes with which object, and hence leaves out

information included in the original statement. One cannot infer the original statement

from the alleged translation of it.

Second, we could *iterate* the adverbial modifiers: "S senses redly, squarely; and S

senses bluely, circularly." The second adverb modifies the first. But there are two

problems with this strategy. First of all, it seems arbitrary to use color-adverbs to modify

S's sensing, reserving shape-adverbs to modify S's sensing-coloredly. Why not the other

way around? The fact that the strategy *requires* a choice here, yet there doesn't seem to

be any reason to choose one way over the other, is a sign that there's something wrong

with the strategy. And there does seem to be a deeper problem with it. Compare the

following case. "S feels righteous anger." We could translate this ascription as "S feels

angrily, righteously." And in this case the iterating of adverbs is correct, because the

second adverb serves to specify the determinate type of angry feeling in question—

namely, a righteous type. Similarly, we could translate "S senses bright red" as "S senses

redly, brightly," understanding the second adverb as a precisification of the first. But the

relationship between "red" and "square" in "S senses a red square" is different from the

relationship between "red" and "bright" in "S senses bright red": while bright red is a determinate of red, red square is not a determinate of red (or of square); these are just independent features, both of which are presented in S's experience. Hence the adverb-iteration strategy conflates cases in which a modifier is being used to precisify another modifier, with cases in which two modifiers are being independently predicated.[2]

Finally, we could *combine* the adverbial modifiers into a simple predicates: "S senses redly-and-squarely and bluely-and-circularly." We can discern the trouble with this strategy by observing that if we know that S sees a red square, we can infer that S sees something red. But we couldn't do that if 'redly-and-squarely' were understood as a simple predicate.

In short: adverbialism seems to be the way forward if we want to square monadicism about intentionality with the way we talk about intentional states, yet adverbialism seems to lack the logical resources to capture the structure of intentionality. Now, there may be ways to add to those logical resources. For example, Tye (1984) has suggested that adverbialists invoke a new operator, which he calls 'Coin' (for "coincidence"). The Coin operator does not merely *conjoin* adverbial modifiers but *links* them in a way that signals the intimate relationship between some but not others. According Tye's proposal, we should analyze "S senses a red square and a blue circle" as "S senses [redly Coin squarely] and S senses [bluely Coin circularly]." We can thereby

---

[2] There is another way to read 'S senses redly, squarely' that Jackson does not discuss. We could understand 'squarely' as a modifier not of 'redly' but of the verb-phrase 'senses redly'. The grammar itself doesn't make it clear exactly what it would mean for a sensing-redly to be squareish rather than for the reddish manner of sensing to be squarish. One way to make sense of it would be if the sensing-redly is a concrete event, in which case the suggestion is that while the *subject* has the feature of undergoing a reddish-sensing, *her undergoing a reddish-sensing* has the feature of being squarish. If so saying makes any sense at all, it doesn't seem to make the right sort of sense: it doesn't correctly capture the structure of the experience we're trying to describe.

get the inferential profile of sensation-sentences right: we can just stipulate that it does *not* follow that S senses redly Coin circularly, but it *does* follow that S senses redly.

Tye's strategy is fine as far as it goes. But it's important to see that even if Tye has shown how to make the grammar of adverbialism adequate to preserve our intuitions about intentional states, the very same concerns crop up again at the level of the metaphysics of monadicism.[3] In short, Tye's strategy is successful only if the way that his operator links modifiers is reflected in the way that some relation links properties within mental states. The problem is that while there are a number of ways for an object to house multiple properties, it is hard to see how any of those ways yields the right sort of linkage between them.

To see how Jackson's problem for adverbialism amounts to a problem for the metaphysics of phenomenal-intentional states, it will helpful to draw a distinction between two ways that a feature can be presented in consciousness.[4] First, a feature can be *instantiated* by the subject. This is how somatosensory properties (or S-S properties), such as visual, auditory, tactile, proprioceptive and affective sensations, are presented to the subject in consciousness. For example, when a subject feels pain, its painful quality is present to her because she instantiates the phenomenal property *painfulness*. But there is another way for a feature to be present to a subject in consciousness, viz. when it is the intentional object of a P-I property. For example, if a subject instantiates the P-I property whose intentional object is *causation* (or "P-I causation" for short), then causation is

---

[3] Tye has subsequently given up on adverbialism; he now holds a representationalist theory of mental states. And his reason for the change seems to have to do with the inability of providing a metaphysical underpinning for adverbialism. See Tye (2009), ch. 5.

[4] To prevent confusion, in the present discussion I restrict the term 'property' to phenomenal properties, while using the term 'feature' more broadly to include both phenomenal properties and the intentional objects of phenomenal-intentional properties.

presented to her in consciousness, though not of course by being instantiated, but rather by being *intended* (that is, by being an intentional object to which her experience points). We can pictorially represent the distinction in the following way:
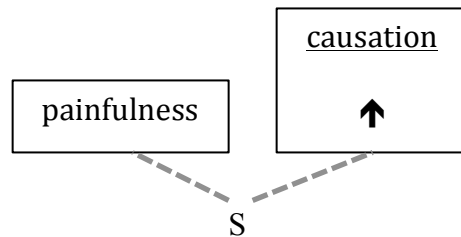

A subject, S, is here represented as instantiating two phenomenal properties: S-S painfulness and P-I causation.

 I understand the example I have been subjecting to Jackson's critiques as follows. For a subject to sense a red square and blue circle requires that the subject be presented with four features: redness, squarehood, blueness, and circularity. Are these features instantiated or intended? The answer is: either, though we will not get around to exploring how phenomenal colors can be intended, or how shape-features can be instantiated, until the next chapter.[5] So let's treat the case as one where redness and blueness are instantiated—call these phenomenal properties "S-S redness" and "S-S blueness"; while squarehood and circularity are intended—call these properties "P-I squarehood" and "P-I circularity."

---

[5] See chapter 5 section 3. My position is that shape-features cannot instantiated, strictly speaking, though there is a type of instantiated feature that is very much like a shape, viz.: the occupying of a portion of the visual field.
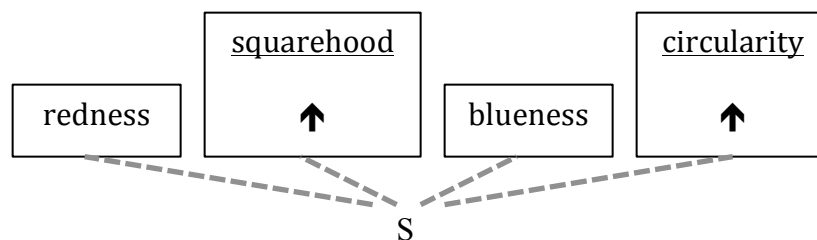
Here's the problem: simply by instantiating these four properties, the subject does not thereby sense a red square and a blue circle; the features they present are not adequately connected. It certainly doesn't make sense to *iterate* them; it is not as though (this instance of) P-I squarehood, say, itself instantiates any phenomenal properties. Nor does it make sense to *combine* the properties. That is, we might say that the subject is presented not with four properties but with two: red-squarehood and blue-circularity. So saying would commit us to denying that the subject's experience has anything in common with an experience as of a red circle and a blue square. (Moreover, it isn't clear how instantiated features could combine with intended features. That would be like combining the color of printer ink with the referents of the words printed with that ink.) In sum: the logical problem Jackson poses for adverbialism can easily be adapted as a metaphysical problem for phenomenal intentionality, and the problem comes to the apparent lack of anything that could play the role of conferring semantic structure on the features presented to a subject in consciousness.

So adapted, Jackson's problem is reminiscent of an older and more general problem in philosophical semantics, the so-called "unity-of-the-proposition" problem. Propositions are abstract entities that philosophers of mind and language appeal to in order to make sense of, and to draw connections between, a number of semantic phenomena. Propositions are variously understood as the semantic content of sentences

146

and/or utterances; as the bearers of truth-values; as the objects of psychological attitudes such as belief; as values of propositional variables in logic, of "-ism's" in academese, and of demonstratives in common parlance ("You don't really believe *that*, do you?").

Among those who believe in propositions, there are roughly three families of views regarding their nature. According to the first, propositions are sets of possible worlds. For example, on this view the proposition that Fluffy is a quadruped is the set of all possible worlds in which Fluffy exists and is a quadruped. This view suffers from famous and probably fatal flaws: it seems to entail that all necessarily false propositions are identical (because identical to the empty set), and all necessarily true propositions are identical (because identical to the set of all possible worlds). I'll have no more to say about the possible-worlds view of propositions.

According to the second, propositions are unstructured, intrinsically representational abstracta. On this view, the proposition that Fluffy is a quadruped is an abstract entity that intrinsically represents the state of affairs of Fluffy's being a quadruped, and is true if and only if and because Fluffy is a quadruped. Call this "the unstructured view."

According to the third, propositions are *structures*, either of individuals and properties, or of abstracta that intrinsically represent individuals and properties (e.g., concepts or senses or modes of presentation). On this view, the proposition that Fluffy is a quadruped is a structure that intrinsically represents the state of affairs of Fluffy's being a quadruped, in part by having as constituents (a) Fluffy and quadrupedality, or else (b) the singular concept of Fluffy and the concept of quadrupedality. Call this "the structured view".

The unity-of-the-proposition problem, made famous by Bertrand Russell but which has ancestors in Plato and Aristotle,[6] is the problem of figuring out what differentiates structured propositions from mere aggregates of sub-propositional constituents. One natural suggestion is to say that the relation that holds between (e.g.) Fluffy and quadripedality—viz., *having* or *instantiating*—must also be included or represented among the proposition's elements. But so including just adds one more item to the aggregate. Note the affinity with Jackson's problem: just as a conjunction of adverbial modifiers fails to capture the difference between a unified percept and a mere plurality of sensations, so a set of propositional constituents fails to differentiate between a unified proposition and a mere plurality of elements.

If there is no way to account for the unity of structured propositions, it seems that we should go in for the *unstructured* view. But the unstructured view is subject to a still more damning objection. If propositions have no constituents, then there is nothing in common between two propositions that for all the world seem to be about the same thing. If the proposition that Fluffy is a dog is an unstructured primitive, it has nothing more in common with the proposition that Fluffy is a quadruped, than with the proposition that there are no unicorns, and that seems wrong.[7] Again, note the affinity with Jackson's problem: just as a combined adverbial modifier fails to capture the similarity-relations that can hold between distinct, complex percepts, so the unstructured view fails to capture the similarity-relations that hold between distinct propositions.

---

[6] Jespersen (2012) nicely summarizes the problem's history.

[7] By stipulation, *that Fluffy is a dog* and *that Fluffy is a quadruped* do have something in common, viz.: their truth-conditions include Fluffy's being some way. But there is no accounting for their sharing this feature; there is nothing *in virtue of which* their truth-conditions have something to do with Fluffy.

A close analog of the unity-of-the-proposition problem can be posed for the metaphysics of phenomenal-intentional states. Suppose a subject entertains the thought, *circularity is beautiful*,[8] and that she does so in part by instantiating P-I circularity and P-I beauty. By instantiating these two properties, she has merely been presented with circularity and with beauty, not (yet) with circularity's being beautiful. We could add that the *having*-relation is intended in her experience as well; but the addition does nothing to connect P-I circularity and P-I beauty in the right way. Again, the problem comes to the apparent lack of anything that could play the role of conferring semantic structure on the features presented to a subject in consciousness.

Recently several philosophers have tried to solve the unity-of-the-proposition problem precisely by first solving the problem of how intentional mental states can have semantic structure, and then by identifying propositions with mental types. For example, Scott Soames (2010) identifies propositions with event-types of special sort, viz., one represented items' being predicated of another in a mental state. Peter Hanks (2011) and Bjorn Jespersen (2012) develop closely related views, Hanks in terms of act-types and Jespersen in terms of procedure-types. Now, the question of whether an account of the semantic structure of mental states can provide us with a solution to the unity-of-the-proposition problem is not my concern here.[9] But what is of concern is whether what

---

[8] A contrived example, to be sure, but nevertheless a thought that many have had. John O'Donohue writes: "The circle is one of the oldest and most powerful symbols. The world is a circle; the sun and moon are too. Even time itself has a circular nature; the day and the year build to a circle. ... The circle never gives itself completely to the eye or to the mind but offers a trusting hospitality to that which is complex and mysterious; it embraces depth and height together" (*Anam Cara*, p. 79).

[9] I reject the view that propositions are mental types of any sort. This is not just because it is strikes me as a category-error to suggest that propositions can be tokened, instantiated, performed or implemented. (The weirdness is not lost on its advocates; Soames (2010), Hanks (2011) and Pitt (2009) all discuss this objection.) The deeper problem is that, while all intentional mental states are in fact phenomenal states or else dispositions to generate phenomenal states, propositions are not types of phenomenal state. They are, rather, the intentional objects toward which phenomenal state of certain types are directed.

these philosophers say about the structure of mental states can shed light on phenomenal-intentional structure. Here, for example, is Soames:

> The view I will outline locates meaning in thought, perception, and the cognitive acts of agents. ... I start with perception.... One who sees an object x *as red* and tastes it *as sweet* thereby predicates redness and sweetness of it, just as one who feels an object *as vibrating* and hears it *as humming* predicates these properties of it. As a result, the perceptual experience of the first represents x as *being red and sweet*, while that of the second represents y *as vibrating and humming*. In virtue of this, the first agent bears a propositional attitude to the proposition *that x is red and sweet*, while the second bears a similar attitude to the proposition *that y is vibrating and humming*.[10]

Soames is suggesting that what subjects need in order to entertain mental states with semantic structure, in addition to entertaining thought-constituents, is for subjects *to predicate* one thought-constituent of another. Among the mental occurences are *predicatings*, which confer semantic structure on the elements of thoughts. Soames proceeds to extend this idea to account for a number of types of semantic complexity that propositions can have, in addition to atomic subject/predicate structure: truth-functional complexity as well as universal, existential and modal quantification.

The suggestion has promise for illuminating the structure of phenomenal-intentional states. First of all, Soames is suggesting that positing one type of cognitive occurrence, *predication*, can explain the semantic structure of both perceptual states and cognitive states. The suggestion thus signals the possibility of simultaneously solving Jackson's problem and the unity-of-the-proposition problem (or rather, those corollaries of these problems that make trouble for phenomenal-intentional states). Second, Soames is suggesting that predication explains not just the semantic structure of atomic thoughts, but that it is the source of the *generativity* of cognition, in many (if not all) of the ways that cognition can be generative.

---

[10] Soames (2010), p. 7-8.

Still, Soames doesn't tell us anything about what predicating is. And without any further story, one might reasonably worry whether any progress has been made. *Of course*," we might say, once you've got *predicating* in the picture, the problems evaporate; but the question is what predicating could come to.[11] In the next section I attempt to offer a more robust account of semantic structure than Soames' rather meager offering. (At the same time, I suspect that the unity-of-the-proposition problem is unsolvable without invoking some sort of primitive, unity-conferring element. So my account may strike some readers as not all that much more robust than Soames'.)

3. Phenomenal Binding

Suppose a subject thinks about circularity and about beauty, and does so by instantiating P-I circularity and P-I beauty. What else is required for her phenomenal state to amount to her thinking that circularity is beautiful? What is needed for something to connect these two P-I properties, such that the connection between the two P-I properties grounds a semantic relation between the intentional objects presented by the two P-I properties. I call this two-tiered connectedness within consciousness "phenomenal binding." In what follows I try to get clearer on what phenomenal binding could be.

First of all, it is clear that phenomenal binding must amount to a genuinely additional mental element; it is not an internal relation between its relata, i.e., a relation whose holding is guaranteed by the existence and nature of its relata. To see this,

---

[11] "What we have is the idea that *something* makes a belief a representation, a slide to thinking that it must be *a making by the person* whose belief it is, and the introduction of a suggestive label, 'predicate'. For all the illumination it provides, we might as well have said that the agent *zegas* the property of loving to Desdemona and Cassio in that order, where we mean by that whatever the agent does or undergoes or whatever is true of the agent that brings it about that he represents Desdemona as loving Cassio: a we know not what we know not why....it seems more to label a problem than to solve it." (Ludwig [2012], p. 895).

consider Gottlob Frege's strategy for solving the unity-of-the-proposition problem. For Frege, atomic propositions are structures that consist of (a) a proper-name and (b) a concept.[12] The sense of a proper name is complete, whereas the sense of a concept is "unsaturated": it is rendered complete when it takes a saturated sense as an argument. Might a strategy along the same lines work for P-I properties? The idea would be that some P-I properties are incomplete in a manner analogous to a concept's being unsaturated. They are, we might say, binding-hungry. An analogy could also be drawn from chemistry: just as some types of atoms have a strong proclivity for bonding whereas others are more stable, some P-I properties have a strong proclivity for latching onto other phenomenal properties, and this is phenomenal binding. There is no additional phenomenal element besides the two relata.

There are two reasons that this suggestion won't work. First, Frege's strategy requires that some constituents of propositions be essentially saturated whereas others are essentially unsaturated, and hence that the same item cannot be the subject of one proposition and the predicate of another. The propositional content under present discussion, "circularity is beautiful", is one in which *circularity* appears as subject. But surely *circularity* can be predicated of things; so also can it be thought of on its own, as a sub-propositional content. This famously weird aspect of Frege's picture is certainly no less weird if we try to apply his strategy to phenomenal-intentional states. A second reason the Fregean suggestion won't work is that it isn't clear that Frege's strategy works in its original context. There is a difference between an aggregate of two items, one saturated and a second unsaturated, vs. a structure in which the first item saturates the second; and something in addition to the two items is needed in order to explain this

---

[12] Frege (1890/1997).

difference. I conclude that even if it were *necessary* for phenomenal binding for one of the relata to be unsaturated (in some sense), that wouldn't be *sufficient*. So phenomenal binding has to involve something more than the two bound items.

A second observation about the relation of phenomenal binding is that it cannot amount to just another P-I property: P-I predication, say. This should be evident from our discussion of the unity-of-the-proposition problem above. Just because a subject is phenomenally presented with circularity, predication, and beauty, it does not follow that beauty is presented as predicated of circularity in the subject's phenomenal state.

Third, phenomenal binding is an asymmetric relation. There is a difference between thinking that circularity is beautiful and that beauty is circular. The binding of one phenomenal property to another has to ground the predication of one intentional object to the other, but not vice versa.

Fourth, phenomenal binding is not merely a relation among phenomenal properties; it must itself be a phenomenal property. If a subject thinks that circularity is beautiful, not only does the proposition she thereby entertains have a determinate direction of predication; but precisely that direction of predication is presented to the subject. When a subject instantiates P-I properties, not only can the intentional contents presented by those properties be semantically related in certain ways and not others, but they can be *presented as* semantically related in certain ways and others.

So far, we've seen that phenomenal binding is a very peculiar phenomenal feature: it is a phenomenally-presented relation that holds asymmetrically between P-I properties, thereby semantically connecting the intentional objects presented by those P-I properties. In fact it is more peculiar still. Suppose a subject is looking at paint swatches,

and spots a red hue that strikes her as beautiful. In this case, beauty is predicated of

redness in her phenomenal state. But what's different about this case is that one of the

relata in the semantic relation of predication is not an *intended* feature, but an *instantiated*

feature. Binding can hold, not just between two P-I properties, but between an S-S and a

P-I property as well.

Here is what I propose. Phenomenal binding occurs when a subject instantiates a

particular phenomenal property, which I'll call "P-I attribution." P-I attribution is a

monadic property, but it intentionally points beyond itself, in two directions, toward

phenomenal features presented in consciousness, either by instantiation or by intention. It

points in one direction, at the recipient of an attribution, and in another direction, at the

subject of an attribution. We might pictorially represent the way that P-I attribution binds

S-S redness and P-I beauty together, such that a subject is presented with the content, *red*

is beautiful, as follows:
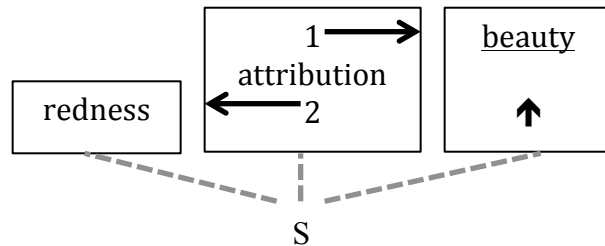
Again, we might pictorially represent the way that P-I attribution binds P-I circularity and

P-I beauty together, such that a subject is presented with the content, *circularity is*

beautiful, as follows:

circularity	1 → attribution 2	beauty
↑		↑

S
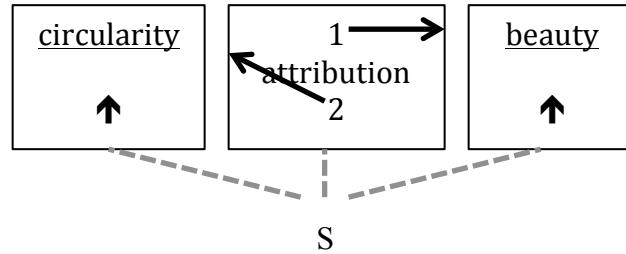
So: phenomenal properties bind when each is pointed to (as it were) by one of the two demonstrations conferred P-I attribution. This is the most basic mechanism whereby propositional unity is presented in consciousness.

Let's return to Jackson's problem to see whether and how P-I attribution can provide a solution. Recall that we framed Jackson's problem in terms of a subject's instantiation of four properties, S-S redness, S-S blueness, P-I squarehood, and P-I circularity. Our question, then, is how these properties are to be bound together such that the subject sees a red square and a blue circle. It is apparent that no amount of binding among these properties is going to deliver the goods. If, for example, S-S redness and P-I squarehood were bound together via an instance of P-I attribution, the resulting phenomenal state would present either the content *redness is square* or the content *squareness is red* (depending on which property is indicated as subject and which as recipient of the attribution). But, of course, if a subject senses a red square, she is not thereby presented with the content that *squareness* is red. It would be more accurate to say that she is presented with the content that *something square* is red. But that's not quite right, either. What is missing is the presentation of *particular instances* of the qualities of squarehood, redhood, etc. How ought we to understand such an aspect of S's

155

experience, an aspect that Michelle Montague has aptly called "the phenomenology of particularity"?[13]

Could we add another P-I property to S's experience, viz., the P-I property whose intentional object is *particularity*? Call this property 'P-I particularity'. This suggestion will not help, and it should be obvious why not. Particularity is a universal: it is the non-qualitative property had by all and only particulars in virtue of being particulars. S's experience does not attribute redness or squarehood to this property; nor does it attribute this property to redness or squarehood. Such attributions are necessarily false, of course (redness is *not* particular, after all: it's a universal). But the more obvious problem is that such constructions have very little to do with the content of perceptual experience, which attributes properties *to particulars*.

Consider an analogous issue in ontology. The world of material objects cannot consist merely in universals. If there are universals—if nominalism about properties is false—then there must be something about reality that particularizes those universals, making it the case that they are instanced. This something—some*things*, really—have been called 'thin particulars' or 'bare particulars' by metaphysicians. I suggest that for S to sense a red square and a blue circle, S's experience must include the *intentional* equivalent of a bare particular. In addition to instantiating P-I properties whose intentional objects are qualities, S must instantiate phenomenal properties that play the role of particularizing those qualities. I'll call these phenomenal elements "P-I units".

Once we have P-I units in the picture, we can see how to solve Jackson's problem: a subject instantiates a P-I unit and other phenomenal properties are bound to the unit via P-I attribution. We can pictorially represent the resulting phenomenal

---

[13] See Montague (2010) for a nice treatment of this issue. My discussion is indebted to hers.

structure as follows (where P-I units are represented via a circle, labeled with an 'U', and where P-I attribution is now represented via a single arrow[14]):
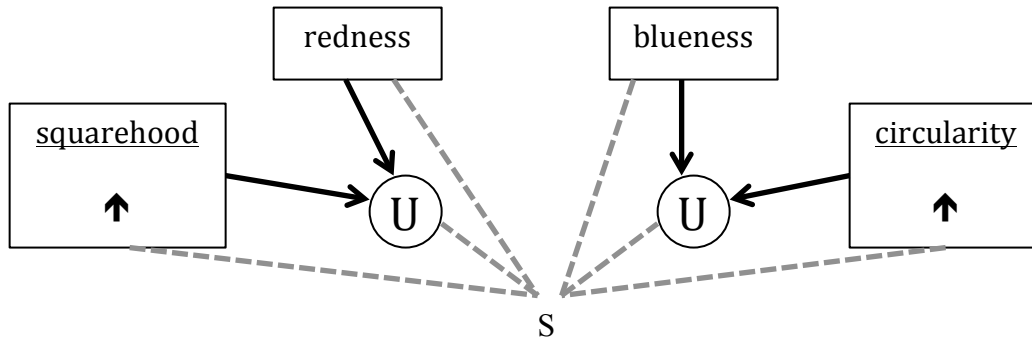
When multiple phenomenal properties are bound up with a P-I unit, I'll call the result a "P-I property cluster." P-I property clusters make possible the presentation in consciousness of two sorts of multiplicities: first, the presentation of multiple objects, and second, the attribution of multiple properties to a single object.

Now, I have invoked P-I property clusters as a way to solve Jackson's problem. But once they are in the picture, they shed light on a number of interesting phenomena related to perception. Here are four:

(1) *Object-based attention*. Perceptual attention is generally understood to be "object-based".[15] That is, when we selectively attend to aspects of our perceptual experience, the units of attention typically are not locations within a subjective space, i.e. parts of a perceptual 'field'; nor are they typically locations within a represented, objective space; nor are they typically the properties presented in perception (e.g. patches of color)—though of course we *can* filter top-down attention through any of these

---

[14] I omit the line representing S's instantiation of P-I attribution.
[15] For a helpful review article, see Scholl (2001).

categories.[16] Rather, the typical units of bottom-up attention are items that are perceived as discreet countables. I suggest that the phenomenology of object-based attention is partly determined by the way that phenomenal properties are bound up with P-I units. Those presented features that are bound up with a P-I unit are perceived as features of an attendable and trackable individual.

(2) *Conflicting intuitions about sensory qualities*. What is a color? There has been a remarkable lack of disagreement among philosophers about this: at one extreme, some hold that a color is a type of phenomenal property (I include myself in this camp). At another extreme, some hold that a color is a type of surface-reflectance profile, or even that it is the disjunction of the physical realizers of such a profile. Philosophers closer to the former camp argue that the essential nature of a color is present in consciousness,[17] and that truths about color are observer-relative.[18] Philosophers closer to the second camp argue that colors only show up in perceptual experience as features attributed to external objects[19]—precisely in contrast to phenomenal properties such as pains. The present picture *predicts* that intuitions would pull in two different directions, when it comes to sensory qualities such as colors. P-I property clusters can involve the attribution, to the very same particular, of *both* instantiated features and intended features. And that's weird: while both are ways for features to be presented in experience, instantiated features are properties of the subject, whereas intended features are precisely *not*: they are the *objects* of phenomenal directedness. It is as though experience confuses

---

[16] Indeed, our ability to direct top-down attention to subregions of expanses within a perceptual field is one key source of quantificational thought. See section 6 below.
[17] See Johnston (1992).
[18] Cohen (2009).
[19] Harman (1990).

the subjective and the objective, thereby committing a projective fallacy (akin to: "it's really tired in here").

(3) *Perceptual vs. cognitive content.* My jumping-off point for our discussion of phenomenal binding was Soames' suggestion that predication is a mental occurrence that explains the semantic structure of both perceptual states and cognitive states. Again, philosophical opinion is sharply divided regarding the relationship between perceptual and cognitive content. On the one hand, perceptual experiences can be assessed for accuracy and inaccuracy, just as belief-states can.[20] On the other hand, it doesn't seem quite right to say that perception is a propositional attitude like belief. For example, David Pitt, who like Soames wants to identify propositions with mental types, nevertheless pushes back against the idea that the relevant mental types can be found both in cognition and in perception:

> The very richness that makes perceptual experience so useful seems to entrain a degree of vagueness in what one might suppose to be their propositional contents... the density and possible propositional indeterminacy of perceptual experiences suggest that, though propositions can be useful in characterizing their contents, and though they might bear interesting relations to propositions, they do not themselves, qua non-conceptual, have the same sort of intentional contents as thoughts.[21]

This is a bad result. If perceptual and cognitive states have different sorts of intentional contents from one another, it isn't clear how we can recognitionally sort them with respect to one another—that is, how introspection can rationally guide our judgments about when a cognitive state and a perceptual state have contents in common.[22] This

---

[20] See discussion in Chapter 2.
[21] Pitt (2009), pp 132-133.
[22] It is noteworthy that Pitt's strong commitment to cognitive phenomenology in his (2004) is motivated by concerns about introspective discrimination, as I discussed in chapter 2. Pitt's comments in his more recent paper indicate that he has not considered the significance of our ability to introspectively compare cognitive and perceptual states for sameness/difference of content.

suggests that whatever the difference between "conceptual" and "non-conceptual" content comes to, this difference should not be overdrawn, thereby rationally isolating perception from cognition. On the other hand, there is surely something to Pitt's observation about the "propositional indeterminacy of perceptual experience." Perceptual experience is not very much like a list of discrete propositions; if we were to try to construct such a list on any given occasion, we might not know where to begin and we certainly would not know where to stop.[23]

P-I property clusters shed light on the senses in which perceptual experience is similar to and different from propositional thought. Suppose a subject has a visual experience as of a red square at some distance from her. On my view, the experience will be structured as follows: she instantiates a P-I unit, which is in turn bound to three phenomenal properties: S-S redness, P-I squareness, and a P-I property as of an egocentric spatial relation, which for now I'll call P-I yonderhood (as in, the red square appears to be located over yonder).[24] We can represent the subjects' experiential state as follows:

---
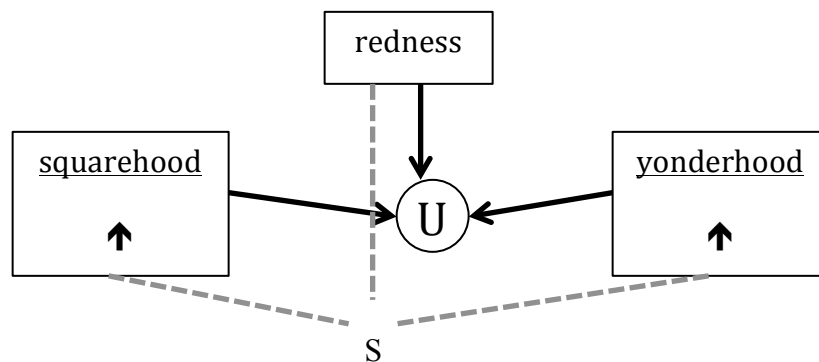
[23] This might be true of cognition, as well. See Siewert (1998), p. 276ff.
[24] Yonderhood is not a property, since it essentially involves an indexical: a thing is yonder if it is a certain distance *from me*. Yonderhood is thus what Andy Egan (2006) calls a "centering feature." Of course there are many determinates of yonderhood, one for every discriminable egocentric distance.

160

What I want to suggest is that the propositional content of such an experiential episode is triply indeterminate. First of all, it is indeterminate between a definite description and a proposition: it would be correct to characterize the content of the experience either as

(1) *the red, square thing yonder,* or as

(2) *There is a red, square thing yonder.*

Second, it is indeterminate between an existentially quantified proposition, a demonstrative proposition, and an atomic proposition: it would be correct to characterize the content of the experience as

(3) *There is a red, square thing yonder*, as

(4) *That is red, square and yonder,* or as

(5) *The red square is yonder.*

Finally, it is indeterminate what is atomically predicated of what: it would be correct to characterize the content of the experience as

(6) The yonder thing is square and red, as

(7) *The red thing is yonder and square*, or as

(8) *The square thing is red and yonder*,

or, indeed, as a proposition that includes only one or two of the three presented properties. Any of these contents can be correctly abstracted from the subject's experience and attributed to her as the contents presented to her in experience. Hence perceptual content is built out of the same elements as cognitive content, without perception's amounting to the sorts of straightforwardly propositional judgments that paradigmatically occur in cognition.

(4) *Demonstrative thought*. Conscious perception makes demonstrative thought possible. I can wonder about *that* tree, marvel at *that* snowman, appreciate *this* guitar solo, and so forth. Such cognitive states cannot be characterized using qualitative language alone. Instead, they require that a referential hook-up with the world has occurred—presumably in perception—and then exploit that referential hook up. I suggest that property clusters can perform this bedrock referential function, at least when four distinct conditions hold.

First, the cluster has to serve as a mode of presentation of a local particular. This occurs when the cluster includes (a) features that present the referent of the cluster as part of the extra-mental world, such egocentric spatial features, and (b) features that serve to discriminate the referent of the cluster from other elements of the perceptual scene, such as shape-features.

Second, the cluster has to include an S-S property. This property grounds the referent of the cluster in vivid, manifest, occurrent reality (rather than in a merely imagined or conceptualized locality).

Third, the cluster has to correspond to an object in the world, such that this object instantiates all or most of the intended features included in the cluster, thereby satisfying the descriptive content the cluster represents. (In cases where no object in the vicinity instantiates all the features bound up with a P-I unit in a perceptual state, reference may be indeterminate[25].)

Fourth, the same object that satisfies the descriptive content of the cluster must also be the cause, in the canonical way, of the instantiation of the cluster. "In the canonical way" is meant to rule out cases of "deviant" causal chains such as the

---

[25] See Montague (2013).

162

following: I seem to see a red ball yonder; there is a red ball yonder; the red ball is emitting gases that cause me to hallucinate that there is a red ball yonder. (I leave it open how precisely to cash out canonical causation, trusting that there is a way to do so.)

Perceptual reference can break down if any of these conditions fail: if the cluster is too descriptively impoverished to single out an object (as may occur in episodes of blurry vision or poor hearing); if the cluster fails to include an S-S property (as may occur in episodes of blindsight); if no object answers to enough of the features presented in the cluster (as may occur in episodes of delusion); or if nothing that answers to enough of the features presented in the cluster is the non-deviant cause of the cluster (as may occur in episodes of hallucination). But if these conditions are satisfied such that reference *has* determinately occurred in a perceptual episode, then this episode can serve as the referential anchor whereby the subject can proceed to think about the object, remember the object, make plans with respect to the object, and so on.[26]

4. Perceptual categorization and cross-modal predication.

I have proposed a mechanism for introducing semantic structure into phenomenal-intentional states, viz., binding via P-I attribution, thereby solving the unity-of-the-proposition problem (or the phenomenological corollary thereof). I have extended the proposal so as to solve Jackson's problem, by proposing phenomenal elements that bring particularity into the intentional picture, viz., P-I units. According to the picture that has emerged, P-I units, and the property clusters that form around them, play a central role in perceptual intentionality: they present complex contents to the subject, from which propositional accuracy-conditions can be abstracted.

---

[26] Much more on this in the chapter 5.

Intuitively, there is more to the intentional structure of perceptual states than P-I property clusters can provide. Here are two examples of perceptual states that appear to have additional structure. (1) *Perceptual categorization*: A subject is looking at a leafy twig. Among the apparent contents of the scene is a green, leaf-shaped thing. Suddenly she realizes that it is a well-camouflaged insect. (2) *Cross-modal binding*: A subject is presented with a persistent, high-pitched sound. At the same time, she sees that a red object on the wall above her is emitting a flashing light. Suddenly she realizes that the sound is coming from the red object on the wall.

In each of these cases, the subjects' experience presumably contains a P-I property cluster. Take the perceptual categorization case: the relevant property cluster includes S-S greenness and P-I leaf-shapedness. In some fashion, P-I insecthood is also included. The question is: in what fashion? The simplest answer to the question would be that it is included in just the way that the other two properties are included. But this doesn't seem to do justice the phenomenology. There is a sense in which insecthood is foregrounded or highlighted in the subject's experience in a way that greenness and leaf-shapedness are not. One way we might illustrate the asymmetry is in thinking about relative aptness of ways of characterizing the accuracy-conditions of her experience. Her experience is accurate only if the insect in front of her is green. But it seems more apt to swap subject and predicate here, i.e. to say that her experience is accurate only if the green thing in front of her is an insect.

We might try to capture the asymmetry by adding that some members in a P-I property clusters can be presented as *more salient* than others. As a first approximation: if a feature is presented in consciousness as salient, the subject takes it as of particular

164

interest or as more attention-worthy than its fellows.[27] I suspect that's part of the story of

why some ways of abstracting accuracy-conditions from perceptual episodes seem more

apt than others. But I doubt it is the whole story. For one thing, perceptual experience is

typically thick with categorization, where most of this categorization is not presented as

salient at all, in the sense of attention-worthiness. (Consider: hearing noises *as words*;

seeing a thing *as a tree*.) Another problem has to do with the way that relational features

are presented in consciousness. For example, suppose it occurs to a subject that the tree

visually presented to her is the very same tree that she planted there when she was a

child. In this case, a P-I property cluster that is part of visual experience is bound with a

complex of phenomenal-intentional properties (in her visual imagination, perhaps) that

presents to her the tree that she remembers. Here we have two clusters of P-I properties,

bound together with the phenomenal equivalent of numerical identity. I will return

presently to the question of how P-I properties as of relational features enter into

phenomenal binding. The important point for present purposes is that this rather elaborate

form of perceptual categorization cannot obviously be understood in terms of the

attributing of more features to a single P-I property cluster. After all, there seem to be two

P-I property clusters at play.

What is needed, I propose, is a phenomenal analogue to parentheses in logic, i.e. a

way for presented contents to be grouped and subordinated, such that presented contents

can be richly interconnected without thereby forming one giant P-I property cluster.

---

[27] What is phenomenal salience? It cannot be a P-I property, the intentional object of which is attributed to
other features presented in experience: otherwise, in the case currently under discussion, the subject's
experience would present the content, *insecthood is salient*. A propositional content that would more
closely capture the content of the experience would be *currently, that thing's being an insect is salient*, i.e.
a temporally-quantified attribution of a property to a property-instance. But even that seems wrong:
salience is not attributed in the same way that presented features are attributed. Salience should be
appropriated to *attentional* rather than *intentional* phenomena.

Indeed, there is independent reason to think that something like this goes on in consciousness. To see this, note that whatever *phenomenal structure* comes to, it must differ from other forms of metaphysical structure. Suppose that a group of particles is arranged chair-wise. It is possible to wonder whether there is a further thing, a chair, in addition to the particles. But now suppose that a group of phenomenal properties is bound together to form a P-I property cluster, e.g. visual experience of a red circle. Is it possible for the subject to wonder whether such properties are *really* bound together? I don't think so, at any rate not without ignoring the manifest nature of her experience. That is, when phenomenal binding occurs, it is not just that a group of phenomenal properties are unified; they are *presented as* unified. So, when a subject instantiates a P-I property cluster, it must include a further phenomenal element, in addition to a P-I unit, instances of P-I attribution, and the various phenomenal properties bound together. All of these elements must be *subsumed* into a phenomenal unity; they must be presented as unified.[28]

What is phenomenal subsumption? It is clear that it cannot amount to the inclusion of a new structural relation that holds among presented features; adding structural relations in no way guarantees that the subject takes the structure as a unity. There is, in other words, an additional unity problem at play in the structure of phenomenal states, besides the phenomenal analogue of the unity-of-the-proposition problem. Just as the addition of new elements to an aggregate does not confer propositional unity on that aggregate, so also the addition of connective tissue to a phenomenal structure does not render the structure *presented as unified* to the subject. So subsumption is something else, and I am inclined to take it as a primitive something else:

---

[28] I borrow the language of subsumption from Bayne & Chalmers (2003). Bayne and Chalmers hold, for phenomenological reasons, that all of a subjects' phenomenal states at a time are presented as unified. I don't see the need for a global subsumption-mechanism, but I do see the need for local subsumptions.

it is a phenomenal property that encompasses phenomenal structures and presents them as unities. It cannot take the place of P-I attribution, since phenomenal subsumption does not *semantically* connect the features presented within a phenomenal structure. Indeed, it does not structure them in any particular way. But given that a structure includes presented features that are semantically connected (via P-I attribution), phenomenal subsumption makes such connectedness manifest to the subject.

Just as parentheses can nestle in a logical formula, so phenomenal subsumptions are able to nestle within a phenomenal structure. Hence, among the features presented in a phenomenal state, some can be presented as more deeply unified with one another than with others—even if all of them are connected via P-I attribution. I suggest that this is what is going on in the cases of perceptual categorization and cross-modal binding introduced above. When a subject sees a green, leaf-shaped thing as an insect, we could pictorially represent her phenomenal state as follows (I henceforth omit a representation of the subject and of instantiation-relations):
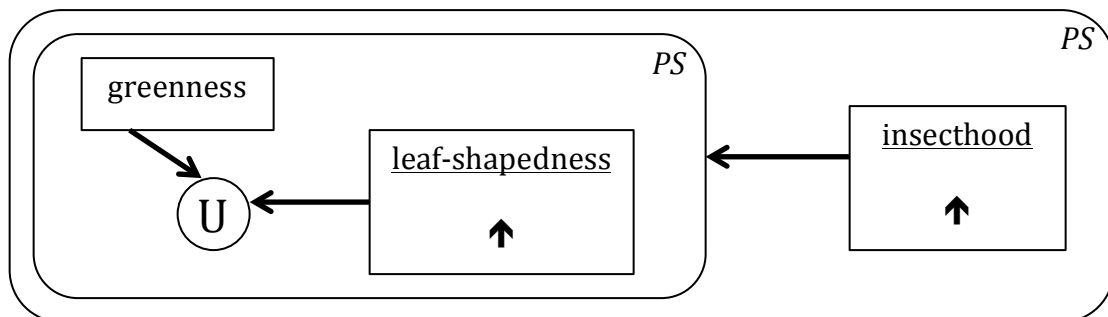
Thus the whole green-leaf-shaped-insect percept is presented as a unity, but parts of its features are treated as an especially unified bundle; and the feature of being an insect is attributed to that bundle. The example of cross-modal binding given above (viz., a subject's attributing a sound to a visually perceived thing) could be depicted analogously.

One lesson of the present discussion is that there is a lot of connectively across phenomenal modes; indeed, when we are talking about the binding of P-I properties to S-S properties, it is not obviously meaningful to categorize the resulting structure as perceptual or cognitive. There is no structural difference between seeing a perceived object as F and judging a perceived object to be F. Now, when it comes to cross-modal binding, there may be a real difference between attributing a visual feature to an auditory object vs. attributing an auditory feature to a visual object—if, for example, the visual features are presented as part of a subsumed P-I property-cluster, and the auditory features are attributed to that subsumed (visual) unity. But P-I units themselves are neutral between perceptual modes; indeed, they are neutral between perception, imagination and cognition (though we have not yet seen them in action in imagination or cognition). In sum: the content of consciousness is more richly interwoven than we may have suspected.

5. Relational binding.

Suppose a subject visually perceives two trees. One appears to her to be farther off than the other. She is asked which is the taller of the two trees, she visually attends; and she reports that tree A appears taller than tree B. How shall we describe the structure of her experience? At the very least, her experience contains two P-I property clusters (one for each tree); in addition, her experience presents the relation of being taller than, which I'll understand in terms of a P-I property, "P-I height-difference". All three of these items are phenomenally bound in some fashion, though the nature of the binding has to look different from what we have seen so far. First of all, P-I height-difference is bound to two

P-I property clusters rather than one. Call this phenomenon "*polyadic binding*". Second, the resulting structure presents to the subject, not just *that* the referents of the P-I property clusters are related by height-difference, but that the relation holds *in a specific direction*. Call this phenomenon "*directionality*". If it weren't for the matter of directionality, the matter of polyadic binding might look unproblematic: we could just say that P-I height-difference is bound, via P-I attribution, to both P-I property clusters. But it is clear that more has to be going on, to explain directionality.

Whence directionality, then? There are four places within the structure of the subject's experience where we might try to locate it: (1) as a distinct, attributed feature; (2) as a feature of the binding that holds between P-I height-difference and the two P-I property clusters; (3) as a feature of the two P-I property-clusters; and (4) as a feature of P-I height-difference. I'll discuss each of these options.

(1) *Directionality as an additional feature*. The idea here would be that P-I height difference is bound to each P-I property cluster via P-I attribution, but a further phenomenal feature—let's call it P-I directionality—somehow affixes to the structure, and indicates to the subject the directionality of the relation. It might look like this:
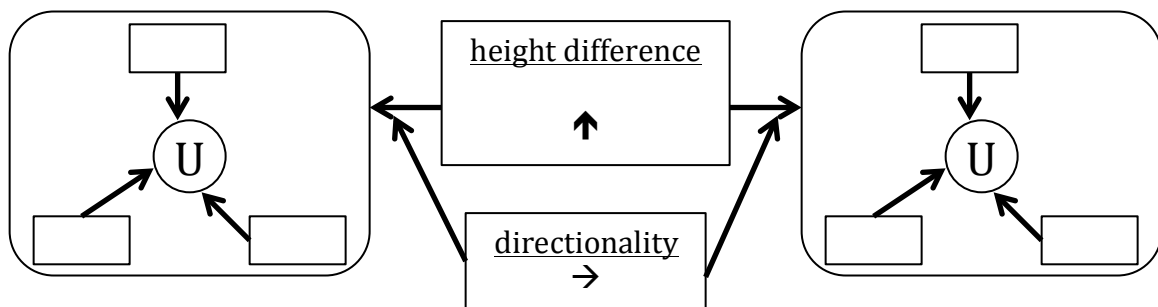

An initial observation is that the method whereby P-I directionality is affixed cannot just be P-I attribution. If it were, then exactly the same question about directionality would

come up for P-I directionality as it does for height difference! That is, the structured modeled above might present to the subject that there is a height difference between the two trees, and it might indicate to the subject that there is directionality to the height difference, but it would not yet indicate that the directionality goes in one direction rather than another. A way to fix this problem would be to treat P-I directionality as akin to P-I attribution, such that it doubly points beyond itself. If these pointings are ordered (as they are for P-I attribution), then the picture above could indeed present to the subject that height difference is attributed *first*, to one P-I property cluster, and *second*, to the other.

But now a deeper problem emerges. Even if we are informed as to the directionality of a relation, or as to the ordering of its attributions, we are not yet in a position to understand the directionality of the relation (somewhat surprisingly, perhaps). Compare: "The causation-relation holds between events A and B, in that order." "The loving-relation holds between individuals X and Y, in that order." It remains unclear what is causing what and who loves whom, because it remains unclear which relatum (cause or effect, lover or beloved), enjoys pride of place in the ordering. The problem seems to be that there is no such thing as generic "directionality" that disambiguates two ways that a relation can hold. Height difference can hold between items A and B either if A is taller than B or if B is taller than A. Simply indicating that height difference is directed from A to B rather than vice versa does not clear up the matter. So the present suggestion isn't adequate.

(2) *Directionality as a feature of phenomenal binding*. Perhaps we should replace P-I attribution with a different mechanism of phenomenal binding, one that uniquely

applies to P-I properties as of relations. Let's call this mechanism "P-I relating." Here is how it might look:

height difference

↑

1
relating
2
3

U

U
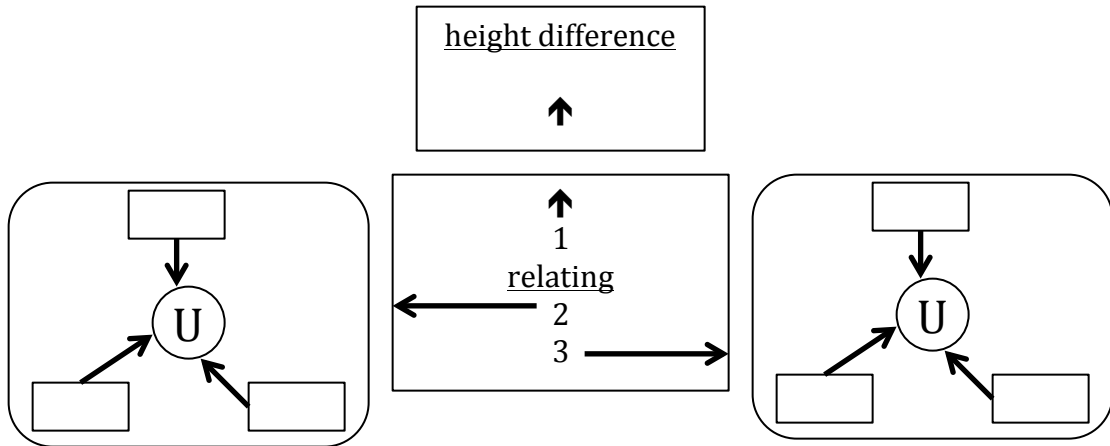
P-I relating thus points beyond itself in three directions: at the relation to be attributed and at each of the relata.

The problem with this option is the same as the problem with option (1). P-I relating presents to the subject an ordering of the connections between the relation and its relata, but nevertheless fails to clear up the directionality of the relation.

(3) *Directionality as a feature of P-I property clusters.* The lesson so far is that directionality is not a generic feature of relational contents. For a phenomenal intentional state to present a relation's holding in one direction rather than another, the particular manner of asymmetry proprietary to that relation must itself be presented in some fashion. Here is one way that might happen: bound to each P-I property cluster is a P-I property that corresponds to one or the other way that a thing can be a relatum of the height-differential relation, viz.: being taller / being shorter. Now, an immediate worry with this view is that a single percept could be presented as taller than one thing and shorter than another. If a single P-I property cluster included *both* P-I tallerhood and P-I

171

shorterhood, the resulting structure would fail to capture the directionality of the height-difference relation. But this problem might be solved if an instance of phenomenal subsumption could select some of the members of a cluster and not others. Here is how we could apply this idea:

tallerhood

shorterhood

height difference

U

U

shorterhood
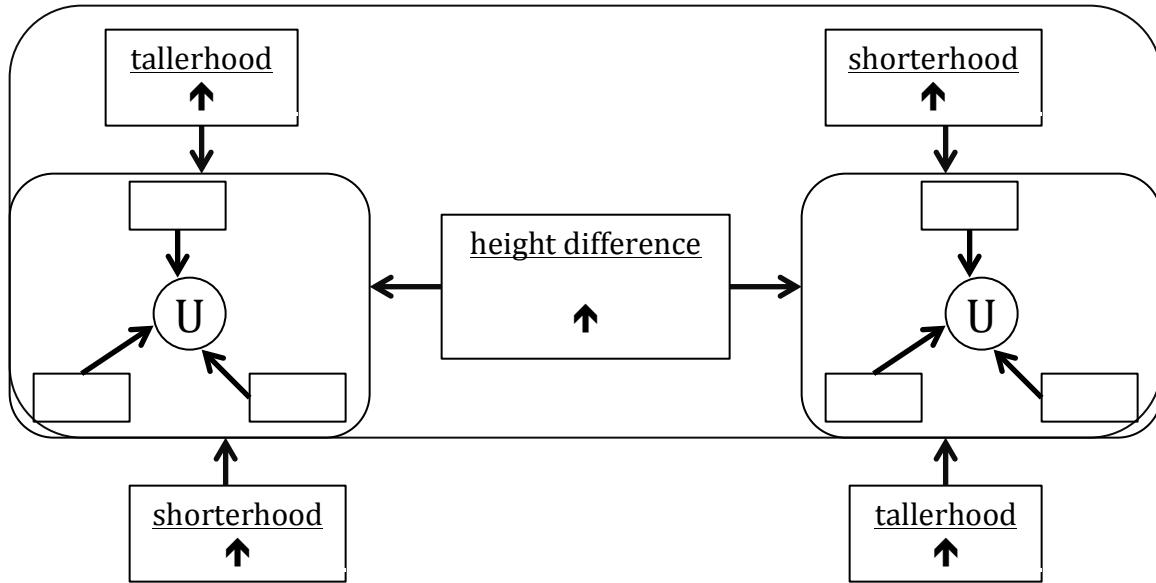
tallerhood

The real problem with this suggestion is that there just is no such feature as 'tallerhood' that can be attributed to the referent of a property cluster. There is such a feature as being taller-than-tree-B. But unless we want to posit such features as primitives (and we shouldn't), it seems to me that the only way for a subject to be presented with A's being than tree B is for the subject to be presented with the holding of a relation between A and B. And that's precisely what we are trying to understand.

(4) *Directionality as a feature of P-I relations.* It seems, then, that the asymmetric nature of relational contents has to be there in the way that the relation itself is presented. I can think of two ways this could be. First, we could think of P-I relations such as P-I height difference as having features analogous to the two terminals of a battery. When

height difference is presented as holding between two trees, one P-I property cluster is bound via an instance of P-I attribution that hooks up to one of the terminals, where the other cluster is bound via an instance of P-I attribution that hooks up to the other, as follows:
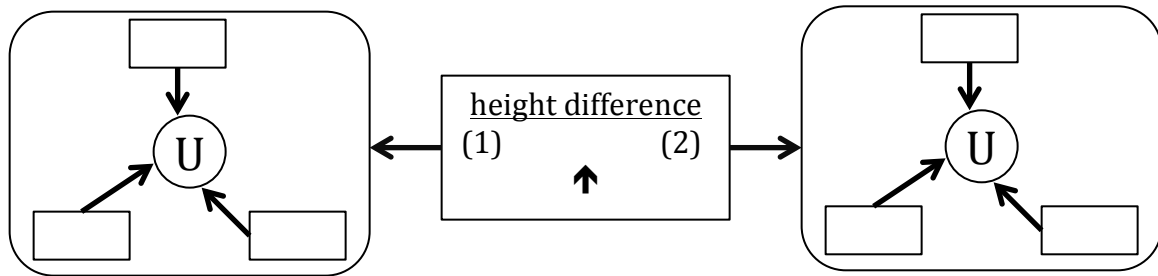

One might initially think that this proposal suffers from the same flaw as did options (1) and (2), viz.: there is nothing in the picture so far that clarifies which terminal is which. Recall, however: P-I properties are not purely syntactic items akin to words. They are, rather, the presentation to the subject of an intentional object—in this case, the relation of height difference. I venture that one who has been phenomenally presented with height difference is one who grasps the sort of thing that height difference is: the relation that holds between two things insofar as and to the extent that one is shorter and the other taller. The two distinct ways for things to be relata in the height-difference relation is given in the nature of the relation itself.

An analogy may be helpful. Consider the symbol we use to denote the greater-than relation in arithmetic: '>'. There are interesting differences between an inscription that makes use of this symbol, e.g. "7 > 4", vs. an inscription that makes use of its equivalent in English, e.g. "Seven is greater than four." Note that the first inscription admits of being read backwards as well as forwards: backwards it translates "four is less

than seven," but of course the English inscription admits of no such reversed reading. The fact that '>' is an iconic symbol allows it to express an asymmetric relationship between two numbers, without privileging either number as the subject of the attribution. The phenomenal-intentional equivalent of '>', which we might call "P-I quantity difference," is neither a lexical nor an iconic representation of quantity difference, but a phenomenal *presenting* of quantity difference. Just as '>' has something like two terminals—its point and its jaw—so it strikes me as plausible that phenomenally-presented relations admit of something analogous.

Still, it might be hard to see what *terminals* could come to in a P-I property. P-I attribution functions by pointing to features presented in experience. So if P-I attribution is going to hook up to presented relations in a way that selects one terminal rather than another, there has to be some way for it to *point to* one terminal rather than another. But if this way of talking is just a metaphor for what actually goes on in consciousness (as surely is), it is hard to see how to trade in the metaphor for something literal. Do presented relations have parts, such that P-I attribution can point to one part rather than another? What sort of parts are they? It is not evident how to make progress here. Fortunately, there is a second way that the asymmetric nature of relational contents could be included in the way that relations are presented. On this alternative, P-I relations point beyond themselves in just the way that P-I attribution does; they are responsible for their own binding with the phenomenal properties that present the relata. Here is the picture that emerges:
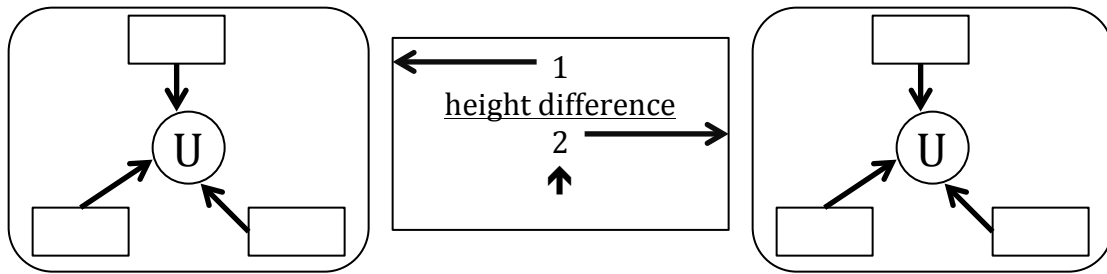
Again, the motivating thought here is that when a relation is phenomenally presented, the brand of asymmetry unique to that relation is likewise presented. And, according to the proposal depicted, nothing further is needed, besides the phenomenal presentation of the relation and the relata, in order for the relation to exhibit directionality.

This suggestion raises an obvious question. If P-I relations can point beyond themselves, thus securing their own binding, why can't all P-I properties do that? Recall that one of the reasons I introduced P-I attribution in the first place was a professed dissatisfaction with the Frege-inspired suggestion that some intentional contents are intrinsically "unsaturated" whereas others are intrinsically "saturated." I have effectively embraced this Frege-esque suggestion when it comes to relations: a P-I relation that fails to point to its relata does exhibit a sort of incompleteness; it points into thin air. And that means that any thought *about* a presented relation—e.g., "height difference is asymmetric"—will involve the attributing of a property to a content that is intrinsically incomplete. The question at issue is why we should embrace such a result in the context of P-I relations but not in the context of P-I properties as of monadic properties.

My answer is that there does seem to be an important difference between the ways relations and monadic features are presented to us. Consider the difference between the property *being red* and the relation *being to the left of*. The former has an intrinsic, qualitative nature that is (in an intuitive but elusive sense) complete in itself—whereas

175

the latter does not; the fact that we express it using an objectless preposition highlights such incompleteness clearly. Now, 'height difference' is a cleaner predicate than 'being to the left of,' but I venture that the item denoted by the former expression is no more qualitatively complete for having a tidier name in English. Attributing properties to relations *just is weirder*, cognitively, than attributing properties to properties.

What this means is that P-I relations share some features with P-I monadic properties and share some features with P-I attribution. They are like P-I attribution inasmuch as they play the role of building complex intentional contents out of simple intentional components. But that does not exhaust their nature, as is the cause with P-I attribution. Like P-I monadic properties, they present unique intentional objects to the subject, intentional objects that correspond to ways things can be.[29]

6. Logical Complexity

So far we have been discussing the structure of phenomenal-intentional states whose content can be expressed using sentences with a subject/predicate form ("redness is beautiful", or a subject/predicate/object form ("Tree A is taller than tree B"). Of course we can understand sentences that are much more complex, for example:

Unless our team scores in the next three minutes, the game is over.

Some spruce trees are over 60 meters tall.

Nothing can be true and false at the same time.

---

[29] Two important addenda to what I have said: (1) P-I relations need not be bound to P-I property clusters; they could be bound to particular P-I properties, as would be the case for a phenomenal-intentional state that expresses the content, "red is similar to orange." (2) Nothing prevents P-I relations from pointing to more than two relata, as would be the case for a phenomenal-intentional state that expresses the content, "Liz gave Sam the lug wrench."

In order to model the structure of such sentences, logicians have introduced logical

operators such as quantifiers, truth-functional connectives, and sentential operators such

as modal and tense operators.[30] The expressive power of formal systems that contain

these elements is indisputable. And given that we are able to understand the sentences

modeled by these formalisms, the elements that make up phenomenal-intentional states

must be at least as expressively powerful. If we were able to identify components of

phenomenal-intentional states that play the same semantic roles as these logical elements,

then we would have gone a long way in understanding how phenomenal-intentional states

can ground arbitrarily complex intentional content. The pressing question is whether the

resources we have already discussed (P-I attribution, P-I units, phenomenal subsumption,

and P-I relations) are adequate to underwrite our understanding of logically complex

sentences, or whether new resources need to be invoked.

Soames (2010) and Hanks (2011) propose a picture according to which no new

structural relations need be introduced, so long as the right sorts of predicates are

cognitively available. Abstracting away from some of the details of and differences

between their accounts, here is what they propose.

Quantifiers: A thinker entertains the proposition 'Something is F' if her cognitive

state predicates *being instantiated*[31] of F-ness; and the proposition 'Everything is F' if her

cognitive state predicates *being universally instantiated*[32] of F-ness. (For our purposes: a

subject instantiates an instance of P-I F-ness, an instance of P-I instantiation / P-I

universal-instantiation, and an instance of P-I attribution that serves to bind them

---

[30] I will have nothing to say in what follows about tense operators. The matter of how representations of time show up in consciousness is simply too big and interesting to treat with any adequacy here; I hope to address the matter directly in future work.

[31] Alternatively: *being true of something*; *being had by something*.

[32] Alternatively: *being true of everything*; *being had by everything*.

together.) A thinker entertains the proposition 'Some As are F' if her cognitive state predicates *being instantiated by As* of F-ness; and the proposition 'All As are F' if her cognitive state predicates *being universally instantiated by As* of F-ness. (For our purposes: a subject instantiates an instance of P-I F-ness, an instance of P-I instantiation-by-As / P-I universal-instantiation-by-As, and an instance of P-I attribution that serves to bind them together.)

Truth-Functional Connectives: A thinker entertains the proposition 'Not-P' if her cognitive state predicates falsehood of P. (For our purposes: a subject instantiates a P-I structure that expresses P, an instance of P-I *falsehood*, and an instance of P-I attribution that serves to bind P-I falsehood to the P-I structure.) A thinker entertains the proposition 'P and Q' if her cognitive state predicates the relational property *conjoint-truth* of P and Q, in that order; the proposition 'P or Q' if her cognitive state predicates the relational property *disjoint-truth* of P and Q, in that order; and the proposition 'If P then Q' if she predicates the relational property *conditional-truth* of P and Q, in that order. (For our purposes: a subject instantiates two P-I structures—one which expresses and the other that expresses Q—and a P-I relation whose intentional object is one of the three logical relations just mentioned, which serves to bind the two structures together.)

Sentential Operators: A thinker entertains the proposition 'Necessarily, P' if her cognitive state predicates *necessity* of P. (For our purposes: a subject instantiates a P-I structure that expresses P, an instance of P-I *necessity*, and an instance of P-I attribution that serves to bind P-I necessity to the P-I structure.) The same things can be said, mutatis mutandis, for 'Possibly, P.'

This framework has a lot to be said for it. It provides a formally adequate theory of several types of logical complexity. While it may require a baroque repertoire of predicates, it certainly preserves parsimony with respect to the combinatorial mechanisms that operate on those predicates. Nevertheless, I have my reservations about its phenomenological adequacy: as a picture of how our phenomenal-intentional states come to express logically complex contents, it is at best incomplete and at worst incorrect.

The framework is incorrect, I contend, regarding the nature of quantification. Intuitively, the thoughts *Socrates is mortal* and *Everyone is mortal* have something in common, viz., they both attribute mortality to something(s). But according to the Soames/Hanks proposal, in the thought *Everyone is mortal*, mortality is attributed to nothing; rather, it is the recipient of an attribution, viz., of *universal instantiation*. I find it incredible both that the direction of attribution is different in each case, and also that the semantic structure of my conscious thoughts could be so unfamiliar to me.

The framework is incomplete in the sense that the predicates it invokes cannot be plausibly treated as primitive P-I contents. By a "primitive P-I content," I mean the sort of feature that is presented in consciousness (by intention). It is plausible that we are presented in consciousness with features such as redness and squarehood, and perhaps many more besides. But it is not plausible that we are presented in consciousness with the feature of *being disjointly true* or *being conditionally true*. Here's why: whatever such a feature comes to, it is not a qualitative nature. These features rather are complicated, formally-specifiable logical relationships. Compare: however it is that I understand the sentence, "Donald served Stephanie with divorce papers," it is not via the instantiation of a primitive P-I relation that presents me with the content, *being served-divorce-papers*.

This is because being served by divorce papers is a complicated phenomenon that involves rich systems of social and legal realities. The sentence encodes no simple qualitative nature but rather a complex relational web. As it is with civil-legal predicates, so it is with logical connectives.

I am concerned for slightly different reasons with the features *being necessary* and *being possible*. There are, of course, many species of necessity and possibility (logical, epistemic, metaphysical, normative). And perhaps there are qualitative natures corresponding to these features. But I do not think, as a matter of fact, that we are presented with them as primitive P-I contents; our grasp of any of these notions inevitably rests on a range of paradigm cases and theoretical considerations, rather than direct, phenomenal apprehension.

In short, insofar as the subject-predicate constructions that Soames and Hanks point to reflect cognitive realities, I do not think they reflect rock-bottom phenomenal realities. In what remains of the chapter I propose to amend and develop the Soames-Hanks view, in hopes that a more phenomenologically plausible picture of the logical structure of thought-contents will emerge.

I begin with quantification. Contra Soames and Hanks, intentional states that present quantificational contents do not amount to attributions of higher-order features—such as *being universally instantiated*—to presented features. Rather, they involve the same intuitive direction of attribution that occurs in non-quantificational intentional states. When a subject thinks, "Some As are F," she does so by instantiating a P-I state in which F-ness is attributed to *some As*. In other words, *some As* can be the intentional object of a P-I structure. Such a structure will contain three elements: (a) P-I A-ness (or a

P-I cluster that denotes A-ness); (b) an element that presents A-ness *as instanced*, i.e. a

phenomenal unit to which P-I A-ness is bound; and (c) a phenomenal element which

serves to delimit the size or amount of instances of A-ness, to *some* of them. This third

element manages to bind to P-I A-ness in some fashion. Such binding is not a matter of

attribution. Rather, such phenomenal-quantificational elements—let's call them

"phenomenal quantifiers"—will provide their own method of binding, just as P-I

relations do. Here is a depiction of a P-I structure that presents *Some As are F*:
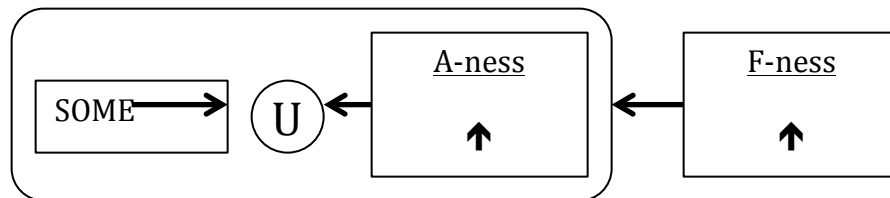
I do not think that phenomenal quantifiers are primitive phenomenal elements;

rather, they are derived from more basic aspects of consciousness. Our most basic

cognitive endowments seem to include two distinct ways of representing quantities in

perception.[33] The first, which Susan Carey calls "parallel individuation," allows us to

home in on *individuals* (up to three or four at a time) and track them through space and

time. This mechanism grounds our ability to count. The second, which Carey calls

"analog magnitude representation," allows us to home in on magnitudes (of size, shape,

etc.) without counting or measuring anything. This mechanism grounds our abilities to

make quantitative estimates and comparative quantitative judgments.

I propose that both of these mechanisms are at bottom *attentional* mechanisms.

This is, perhaps, obvious in the case of parallel individuation: we can *attend* to one, two,

---

[33] See Carey (2009), chapter 4.

three, or four consciously presented individuals at a time. (Call this phenomenon "unit-based attention.") It is perhaps less obvious in the case of analog magnitude representation. But it is nevertheless plausible. We are able not only to attend to individuals, but to expanses within a perceptual scene; further, we are able to attend to *subregions* of expanses within a perceptual scene. For example, I can attend to the expanse of the ceiling above me; I can also attend to the *left half* of the expanse above me, where what is selected by this act of attention is essentially determined by its contrast in magnitude with the expanse of which it appears to be a part. (Call this phenomenon "region-based attention".)

Phenomenal quantifiers are *abstractions out of* these attentional phenomena. Here is the basic idea: what unit-based and region-based attention accomplish is the *selecting* of some portion of the contents presented in a phenomenal-intentional state. What gets abstracted is this selection-function, in its various guises: phenomenal quantifiers that effect "absolute" quantification (*five*, *a few*, *many*, etc.) capture the type of selection-function common to episodes of unit-based attention of certain sorts. Phenomenal quantifiers that effect "relative" quantification (*all*, *some*, *most*, etc.) capture the type of selection-function common to episodes of region-based attention of certain sorts.[34]

Appealing to abstraction raises many questions. There are questions of specifics: how to understand the *inputs* to abstraction, how to understand the *mechanism* of abstraction, and how to understand the *outputs* of abstraction. There are also worries about circularity: since commonality is promiscuous ("everything is like everything else in an infinite number of ways"), it seems that the inputs to an abstraction-operation have to include *which type of commonality is salient*; but abstraction is supposed to *generate*

---

[34] For the distinction between absolute and relative quantification, see Jackendoff (1977), ch. 5.

such an understanding, not *presuppose* it.[35] There is no space here to address these

questions and concerns satisfactorily. For now, I simply observe that, as a matter of

actual fact, we are able to abstract conceptions of types of phenomenal property from

particular phenomenal episodes (sometimes even from a single episode). Think, for

example, of what it is like to feel tipsy, or to feel wired: it seems obvious that I can lack

concepts for these phenomenal properties prior to experiencing them, and that once I do

experience them, I can conceptualize their specific natures unproblematically.[36]

Turning to truth-functional connectives: the basic materials needed are P-I

properties whose intentional objects are *truth* and *falsehood*, and a P-I relation whose

intentional object is *implication*. A P-I state that presents a negated proposition is a state

that includes an instance of P-I falsehood that is bound to a structure that presents a

propositional content. A P-I state that presents a conditional proposition is a state that

includes two proposition-presenting structures that are bound together via an instance of

P-I implication. So far, this is not much of a departure from Soames and Hanks.

Conjunctive propositions can be entertained via *less* machinery than Soames and Hanks

invoke, whereas disjunctive propositions require a bit more. Conjunctive propositions can

be entertained, that is, via the subsumption of two proposition-presenting P-I structures;

no additional attribution of truth is required.[37] Disjunctive propositions, on the other

hand, can only be entertained via the inclusion of a phenomenal quantifier. To entertain

---

[35] See Laurence & Margolis (2012) for clear discussion of these matters.
[36] I say a bit more about abstraction in sections 3 and 6 of the chapter 5.
[37] My discussion of Jackson's problem above assumed as much. I made no mention of subsumption, either. Here's why: were I to characterize all of the intentional contents of your mind at a time, I could do so by making listing a bunch of propositions or by creating a long, conjunctive proposition. Neither would be inaccurate. Now, there might be a sense in which some of the contents you are entertaining are more unified with one another than they are with others, for example because they are subsumed. The easiest way to represent this difference would be to conjoin some of the items on the list.

the proposition *P or Q*, that is, is to entertain the proposition *At least one of {P, Q} is true*. We can pictorially represent such a state as follows:
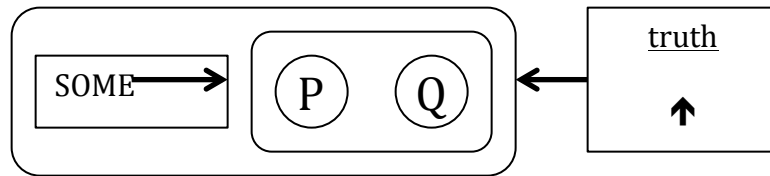
(Or course, a similar P-I structure could present the proposition *P and Q*, if the phenomenal quantifier that expresses *some* were replaced with the phenomenal quantifier that expresses *all*. But such a way of entertaining conjunctions is unnecessarily complicated.)

As with quantifiers, I do not believe that P-I properties whose intentional objects are truth, falsehood, and implication are phenomenal primitives. They are abstractions, not out of attentional phenomenology, but rather out of the phenomenology of epistemic attitudes. Suppose a subject instantiates a P-I structure that presents to the subject the propositional content that P. So far, this is a mere entertaining of P. But of course P could strike her as certain, as likely, as plausible, as implausible, or as certainly false. This "striking" is part of her conscious state; call it the "credence halo" that accompanies the P-I structure that presents her with the content that P. A necessary condition on a subject's consciously judging that P is that the subject instantiates a P-I structure that presents the content that P, and that the structure is accompanied by a sufficiently strong credence halo.[38] It should be clear that credence halos are not bound P-I properties. When a subject judges that P, she is not representing P as being true. This is because she could represent P as being true when she in fact disbelieves P. Credence halos thus amount to a

---

[38] Whether this is a sufficient condition for the subject's judging that P turns on whether judging is a type of action or at any rate is essentially caused in certain ways rather than others.

new type of phenomenal property, a type that is irreducibly epistemic. But suppose that we were able not merely to instantiate these epistemic-phenomenal properties, but to *intend* them, and to attribute such intended features to propositions. The resulting structure would present the content: *P has the property of being what's it's like when P seems credible (/incredible).* The resulting structure would amount to a non-committal attribution of belief-worthiness. I suggest that such an abstracted P-I property serves as the mode of presentation of our concept of truth, and a corresponding abstracted P-I property serves as the mode of presentation of our concept of falsehood. (We then go on to argue about which properties *being true* and *being false* are.)[39]

Credence halos are not the only type of epistemic-phenomenal properties. It can seem to subjects, not just that phenomenal-intentional contents are true or false, but that they bear epistemic relations to one another. Such seemings underly our sense, in thought, of what follows from what.[40] I'll call such seemings "inference-links." Again: inference-links amount to phenomenal features that are irreducibly epistemic. But suppose that we were able not merely to instantiate them, but also intend them, and to connect presented propositions via such intended relations. The resulting structure would present the content: *P's connection to Q is what it's like when P seems to follow from Q.* I suggest that such abstracted P-I relations serve as modes of presentation of our concept of implication. (We then go on to argue about which relation implication is.)

---

[39] Alternatively, there could be one more (objectivizing) step in the abstraction-operation, such that our mode of presentation of truth amounts to *the property P has when P seems credible.*

[40] Anders Nes (forthcoming) has recently suggested that this epistemic connective tissue in consciousness is best understood in terms of Gricean natural (i.e. nonconventional) meaning. Certainly we *express* inferences in those terms; we might say (to use one of Nes's examples), "Tom went to the museum. All who went to the museum will be late. This means that Tom will be late." The virtue of this proposal is that it provides an elegant, non-technical way to describe a very common but nevertheless neglected feature of consciousness. On the other hand, I'm not sure that the appeal to Grice sheds any new light on the phenomenon.

185

Finally, I am quite happy to endorse the Soames-Hanks picture of the cognitive correlates of sentential operators as features, such as being necessary and being possible, that are attributed to propositions. But again, I do not think that P-I properties that present the features of being necessary and being possible are phenomenologically basic. And it turns out that we already have in place all the phenomenal elements from which they are derived. Modal features are, I contend, abstractions out of *inferential* episodes. To suppose that P is to attribute truth to P (in the manner discussed above, viz. attributing *what it's like when P seems credible* to P). Supposing that P thus differs both from judging that P, where P is accompanied by a high-credence halo, and mere entertaining that P, where P neither seems true nor is presented as being true. Inference-links can be presented in consciousness as holding between two propositions P and Q, whether or not either P or Q is accompanied by a high-credence halo. For Q to seem necessary is for Q to be accompanied by a high-credence halo, on any supposition whatever; there is no P such that an inference-link is presented as holding between P and the negation of Q, in other words. For Q to seem possible is for there to be no P such that (a) P is accompanied by a high-credence halo and (b) an inference-link is presented as holding between P and the negation of Q. (More simply: for something to seem necessary is for it to seem consistent with everything; for something to seem possible is for it to seem consistent with everything one believes.) The P-I properties that are abstracted from these episodes serve as modes of presentation of our concept of necessity and possibility. (We then go on to argue about which properties they denote.)

To recap: in order to be able to entertain logically complex propositions, we need at least to be able to instantiate the phenomenal equivalents of quantifiers, connectives

and modal operators. Phenomenal quantifiers are abstractions out of attentional phemenology; connectives and modal operators are abstractions out of epistemic phenomenology. Now, it's possible that at certain stages of cognitive development, we come to grasp quantification, truth, falsity, implication, necessity and possibility quite directly, by instantiating *new, primitive P-I properties* that present these logical notions. (The possibility of such new primitives will crop again in the next two chapters.) In that case we would be able to kick out the attentional and epistemic-phenomenological ladders we used to get there. But I don't think that this *must* happen in order for logically complex thought to be possible.

7. Conclusion.

Let's return to the observations that motivated my speculations in this chapter. Monadicism about phenomenal properties holds that phenomenal properties are intrinsic properties of the subject, rather than relations between the subject and something else. If monadicism is true about phenomenal properties generally, it must be true about phenomenal-intentional properties specifically. But the semantic structure of intentional states poses a prima facie challenge to monadicism. This is because the various ways we know of for multiple intrinsic properties of a thing to be related to each other are inadequate to capture obvious facts about semantic structure. Such properties can be instantiated (1) *conjointly*, where a thing instantiates X and Y simultaneously; (2) *iteratively*, where a thing instantiates X, and X instantiates Y; or perhaps (3) *simply*, where a thing instantiates a new property that somehow merges X and Y. But if P-I properties were related to each other only in one of these three ways, no semantic

structure would emerge: (1) would leave them too isolated from one another, (3) would join them too tightly to one another, and (2) just seems not only unhelpful but unintelligible. (How could my being presented with F-ness itself be presented with G-ness?)

We could conclude from these observations that monadicism about intentional states is false: intentional states must involve relations between the subject and an item that already has the right sort of structure, such as a proposition, a sense datum, or a state of affairs. But we could also conclude that we do not yet have all of the requisite phenomenal elements in view: we were working with the bricks but not with the mortar, so to speak. The present chapter has explored prospects for positing new, structure-conferring phenomenal elements. Central to my proposal is the idea of *phenomenal binding*. This occurs is when one phenomenal-intentional element points beyond itself to another phenomenal element. P-I attribution is the most ubiquitous source of phenomenal binding, though all P-I relations (that is, P-I properties whose intentional objects are polyadic properties) generate phenomenal binding is just the same way. Once we recognize the need for P-I properties that serve to particularize intentional contents (instantiations of which I called 'P-I units'), we are in a position to explain a wide range of data related to the structure of conscious intentional states, including: object-based attention; perceptual demonstration; perceptual categorization; cross-modal predication; relational predication; and the relationship between perceptual and propositional content. Now, these resources aren't quite adequate to explain *logical* structure, where such structure involves the phenomenal equivalent of quantifiers, truth-functional connectives, and sentential operators. But if we suppose there is some mechanism for phenomenal

abstraction, and if we suppose that attentional and epistemic phenomenology can serve as inputs to this mechanism, then we can see how to extend our story to cover logical structure as well.

5. Abstracting and Constructing Contents

1. Introduction.

I have said, in chapter 3, that the mechanism of the phenomenal grounding of intentional contents is as follows: P-I properties combine, in perception, imagination and cognition, to form modes of presentation of all the intentional contents we can entertain. In the previous chapter I discussed a number of ways that P-I properties can combine to form complex structures, including structures that present propositional contents. Now, for some possible cognizer who can instantiate a proprietary P-I property corresponding to every sub-propositional content she can entertain, what I have said so far is enough to explain how phenomenality could ground all of her mental contents.[1] But I am interested in accounting for how P-I properties combine to form modes of presentation of all the contents *we* entertain. And our intentional capacities extend beyond our *phenomenal-intentional* capacities: it is frequently the case that we *come to discover* the nature of a kind or a property that we have long been perceiving, imagining, or thinking about. In Russell's language, *description*, rather than mere *acquaintance*, mediates the vast majority of our referential successes; the contents with which we are acquainted form of a sparse set. In this chapter I venture some empirically-informed guesses about which contents are in the set, and then explain how these primitives can be used to construct our entire intentional repertoire.

---

[1] If, like us, she can entertain indexical contents, then she will also need a way to deploy indices such as 'now' and 'I' and 'here' within phenomenal-intentional states. I will have nothing further to say about indexical contents; I am happy to include indices as primitives if need be, though I hold out hope for a phenomenal-descriptivist account of indices along the lines of Howell (2006).

Thus the project of the present chapter is continuous with a long tradition of trying to identify the basic building blocks of human thought. The project goes back at least to Locke's conceptual empiricism, followed a century later by Kant's conceptual nativism. Carnap famously attempted a similar project in the 20th century in *The Logical Construction of the World*, though of course constrained by somewhat different commitments from those of either Locke or Kant. More recently, cognitive scientists have proposed a variety of theories of the structure of the human conceptual system, providing as evidence both experimental data and common-sense observations of the way we think and speak. My discussion draws inspiration from all of these projects.

A couple of caveats are in order. First, my discussion is necessarily tentative. There is as much disagreement as consensus among cognitive scientists about the nature and sources of our conceptual apparatus. I have had to make choices about which empirical hypotheses to take on as adequacy-constraints and which not to. Second, my discussion will be misleadingly linear and in places comically brief. Mature human intentional capacities involve complex iterations and scaffoldings, the details of which I will only gesture toward.

Here is the plan for the chapter. In section 1, I discuss three methods for determining the phenomenal-intentional primitives and venture a short list of candidates on which the three methods appear to converge. In section 2, I discuss a tier of P-I contents that are in some sense derivative off of those adduced in section 1, but are still among those contents directly grasped by subject. In sections 3 and 4 I move on to a discussion of how the P-I properties in the first two tiers form modes of presentation of wide contents, both of individuals (section 3) and kinds (section 4). Finally, in section 5, I

consider reasons for thinking that new P-I primitives emerge out of the hierarchy I have been sketching. Along the way I propose a theory of what differentiates perceptual, imaginative, and cognitive states.

2. First tier: Primitive P-I properties.

Which intentional contents are primitively presented in our phenomenal states? There are at least three different methodological routes to finding an answer to this question. First, we can investigate the *neural-functional primitives*: we can use the tools of cognitive neuroscience discover the most basic neural structures that correlate with particular intentional properties in consciousness. Second, we can investigate the *developmental primitives*: we can use the tools of cognitive psychology to discover which sorts of intentional contents show up very early in human development and very widely across human cultures. Third, we can investigate the *analytical primitives*: we can use the tools of conceptual analysis and cognitive linguistics to determine which sorts of contents cannot be analyzed or understood in simpler terms. These three methods can help us converge on the *metaphysical primitives*, i.e. those P-I properties out of which all our intentional states are constructed.

Neural-functional primitives. Recent discoveries in cognitive neuroscience have given us good reason to think that there is an asymmetric dependence relation between the sorts of capacities operative in perception and action, on the one hand, and imagination and cognition, on the other.[2] That is, paradigmatic cases of imagination and cognition seem to depend on the mechanics of perception and action, in a manner and to a degree that is not reciprocated. It turns out that brain mappings for perception (in some

---

[2] See Barsalou (2008) for a recent review of the extensive literature on these findings.

mode, e.g. vision) and imagination (in the same mode) are very closely related. The same is true for motor functions: the same parts of our brains are active when we imagine doing something as when we actually do it. More provocatively, recent studies have shown that these same activation-patterns correlate with *linguistic comprehension*. Many cognitive scientists have used these findings to argue for the "grounded cognition" hypothesis: our cognitive capacities are "grounded" in our perceptual and motor experiences. (This paradigm has also been called the "neo-empiricist" approach to mental content.[3]) This suggests that if primitive intentional contents are presented anywhere, they are presented in perceptual and motor experiences, and that imagination and cognition derive their contents from perceptual experience. This is a substantive, if intuitive, idea, and it will structure much of my discussion in this chapter (though there are reasons to think it ultimately comes up short, as we shall see in section 5).

Cognitive neuroscience may be able to help us discover not only where the primitives are to be found (viz., in perceptual and motor experience) but also what the primitives are: by finding correlations between particular neural structures, on the one hand, and particular representational contents presented in perceptual experiences, on the other. When certain of these structures are active (a) early in cognitive development, (b) consistently across subjects, and (c) in connection with lots of different types of conscious intentional states, we have some reason to think that the representational contents with which they correlate are among the P-I primitives. Now, establishing that a neural structure satisfies these three desiderata is tricky, because neuroscientists disagree about which stages along perceptual processing-pathways in the brain correlate with

---

[3] The paradigm is "empiricist" in the sense of locating the sources of our representational capacities in sense-experience. Whether any such capacities are innate or whether they are all learned is an orthogonal matter.

conscious experience.[4] Thus even if we find that some neural structure is causally responsive to a particular type of perceptual stimulus, it doesn't follow that that perceptual stimulus is represented in perceptual experience. Nevertheless, the right combination of experimental design, phenomenological reflection, and progress in isolating the neural correlates of perceptual consciousness may allow us to home in on a candidate set of primitives. Current good candidates include lines, curves, shapes, orientation and egocentric location, as well as relational features such as motion and color- and luminance-contrast.[5]

Developmental primitives. Debates over whether psychological items such as concepts or beliefs are innate or whether they must be learned go back to the 17th century. (Probably they go back to the *Meno* or earlier.) In recent decades, so-called "nativist" theories have enjoyed something of a resurgence. Eric Margolis and Stephen Laurence frame the current debate between "nativism" and its contemporary rival "empiricism" as follows:

> In their lifetimes, human beings come to possess a wide range of concepts and psychological abilities. ... The question for both nativists and empiricists is where all these psychological traits come from. In particular, what kinds of psychological systems are responsible for their acquisition?
> The empiricist answer posits few distinct types of psychological mechanisms, states, and processes for acquiring psychological traits, and supposes that the same systems of acquisition operate across many psychological domains (e.g., the psychological mechanisms for learning natural language are the same as those for learning about object permanence).... The nativist answer, by contrast, posits many distinct types of psychological mechanisms, states, and processes for acquiring psychological traits, and supposes that different systems of acquisition operate across different psychological domains (e.g., the psychological mechanisms for learning natural language are distinct from those for learning about object permanence).[6]

---

[4] Tononi & Koch (2008).
[5] Cf. e.g. Marr (1982), Zeki (1993), and Baars & Gage (2010).
[6] Margolis & Laurence (2013).

Understood thusly as a claim about the diversity and specificity of basic human psychological mechanisms, Laurence & Margolis contend that there is a strong case to be made for nativism. Susan Carey goes further, claiming that "the question of whether development begins with a stock of merely sensory primitives," or instead with dedicated "innate perceptual input analyzers" that generate richly intentional representations of distal objects prior to any associative learning process, has been "conclusively settled" in favor of nativism. For example, there is extensive evidence that newborns, both human and non-human, represent depth-properties prior to any learning.[7]

One particularly prominent version of nativism comes from Elizabeth Spelke, whose "core cognition" model posits four or five representational systems that are present very early in cognitive development and continue to operate throughout adult life.[8] Much of Spelke's empirical case for her view has come in the form of "looking time" studies. The core assumption of this paradigm is that subjects are prone to look longer at unexpected rather than expected outcomes. (The paradigm is especially helpful for studying infant cognition, since infants do not answer researchers' questions.) Here are the representations Spelke takes to be generated by the core systems: (1) *objects* (individuated according to principles of cohension, continuity and contact), (2) *agents* and *goal-directed activity*, (3) *number*, (4) basic *geometric shapes* and *relations*, and (more tentatively) (5) *social conspecifics* (roughly, the categories *one of us* and *not one of us*). Carey modifies this list slightly, subdividing the *number* system into a counting-

---

[7] Carey (2008), pp 30-32.
[8] See Kinzler & Spelke (2007). Spelke herself seems to understand her view as intermediate between radical nativism and radical empiricism. In light of the Margolis/Laurence conception of the debate, it's clear that there is a whole continuum of options here, so whether a view should be counted as nativist or empiricist turns out to be a matter of perspective and of degree.

system and a magnitude-estimate system[9], and adding *cause* as a primitive, citing studies that support attributions of causal-representations to six-month-olds.[10]

Now, there is conceptual space between the set of primitive P-I properties we instantiate with the set of representational capacities that best explain our surprise-reactions. For one thing, it is not clear to me that surprise-reactions are always best explained in terms of intentional states, let alone *conscious* intentional states. For another, even when surprise-reactions *are* explained in terms of conscious intentional states, I think it is possible for content to be implicitly but not explicitly present.[11] These qualifications notwithstanding, I take the core cognition paradigm in developmental psychology as an excellent starting-point for building a list of candidate P-I primitives.

Analytical primitives. It is interesting to note that Carey's case for nativism is not independent of philosophical reflection on conceptual relationships. She thinks that it is evident upon reflection, prior to any investigation in a lab, that there is no way to built an object-conception out of a flurry of sensations, and hence that an object-conception must be psychologically basic.[12] Similarly, Margolis and Laurence argue that some representational capacities *must be* innate, because there is no way to get general categories out of an abstraction-process whose only inputs are representations of particulars.[13] This suggests that a third—and complementary—way to determine which intentional primitives we entertain is to reverse-engineer our conceptual categories and figure out where they bottom out.

---

[9] See Carey (2009), ch. 4.
[10] Ibid., p. 242.
[11] I discuss the distinction in section 4 below.
[12] Carey (2009), p. 35-36.
[13] Laurence & Margolis (2012), "Abstraction and the Origin of General Ideas."

As long as there have been philosophers, they have been undertaking this sort of project, of course. And while conceptual analysis has been on the defensive for some time now, it still has plenty of practitioners and not a few defenders.[14] But I am more interested in a more third-personal, data-driven method of analysis, as developed by cognitive linguists such as George Lakoff and Robert Langacker. These theorists have turned their attention to peculiarities of language-use: for example to ways that subtle changes in syntax modify meaning, or to the usage of entrenched non-literal or quasi-literal linguistic expressions. They use these data to support inferences about the conceptual tools we bring to bear when we use language. Their aim is to show that otherwise inexplicable quirks of language can be easily understood, given hypotheses about how we cognize (and about how the way we cognize motivates the way we speak).

There is no space here to do justice to this relatively new, creative, and explanatorily powerful research program. The important point for our purposes is this: one aspect of the program has been to identify a set of basic categories that seem to underwrite much of our thinking and speaking. For example, George Lakoff and Mark Johnson analyze many types of utterances as involving metaphorical extensions of somatosensory and motor/agential categories.[15] Such categories include: physical object, substance (in the everyday/chemical sense rather than the Aristotelian sense), container (including relations of inside vs. outside), conduit (i.e., means of physical transmission), orientation (up vs. down and front vs. back), light and dark, warm and cold, person,

---

[14] Cf. Chalmers & Jackson (2001), Ludwig (2013).
[15] Lakoff & Johnson (1980), Lakoff (1999).

edifice, conflict, causation, journey.[16] (They do not pretend their list is exhaustive, or that the items on their list are all equally conceptually basic.)

Robert Langacker (1991) contends that basic grammatical categories such as noun and verb also derive from a few basic, experiential categories: "Discrete physical objects are clearly prototypical for the class of nouns, and their energetic interactions for the class of verbs."[17] Hence all nouns and nominals, on the one hand, and all verbs and verb-phrases, on the other, inherit their meaningfulness and grammaticality via metaphorical connections with our folk-physical understanding of the world. (Thus syntax is grounded, in part at least, in semantics.) In addition to the categories of physical object and "energetic interaction," Langacker thinks we need the categories of process, domain, and region (where regions are proper parts or proper subsets of domains).

I don't claim to be an expert on the details of any of these research programs. Perhaps in time they will be substantially altered or even supplanted. In the meantime, they are at least promising and suggestive when it comes to delineating a set of contents we primitively entertain. A conservative list, compiled on the basis of convergence among the three methodological routes, includes object, agent, cause, motion, egocentric location, and some set of shape- and size-properties.

3. Second tier: Perceptual constancies and imaginative states.

---

[16] Regarding the category *causation*, they write: "[C]ausation is a basic human concept...but this does not mean that it is an undecomposable primitive. We would like to suggest instead that causation is best understood as an experiential gestalt. A proper understanding of causation requires that it be viewed as a cluster of other components" (pp. 69-70). They go on to discuss the other components that need to be present for an experience to count as a prototypical case of causation—it needs to take time, it typically involves spatial contiguity, etc. They thus draw attention to an important possibility, viz. that few if any of our ordinary linguistic predicates express primitive P-I properties, but rather express complex conceptions of categories. This possibility is consistent with there being intentional primitives included in any "experiential gestalt" that serves as a prototype for a category. I'm not sure what the alternative would be.
[17] Langacker (1991), p. 14.

Let's suppose that foregoing list comprises the contents we primitively entertain. Let's

suppose, further, that the neo-empiricist/grounded cognition program (according to which

sensorimotor representational systems provide the raw materials for all human

categorization) is on the right track. These suppositions would mean that the items on the

list comprise the representational elements of our most basic perceptual states: i.e., the

features *intended* in basic perception. If we further include features that are *instantiated*

in basic perception (viz., S-S properties such as phenomenal sounds and colors and so

forth), we then have on the table all of the items out of which our basic perceptual states

are composed—at least when such items are suitably subject to *phenomenal binding* (as

discussed in the last chapter).

These resources constitute the "first tier" in the hierarchy of intentional contents

we can entertain. A second tier takes us only slightly beyond the first. This tier manifests

itself in the form of *perceptual constancy*. The phenomenal-intentional elements that

make perceptual constancies possible, I will argue, are in some sense *derivative* off of the

primitive P-I and S-S properties, but nevertheless amount to another layer of *narrow*

content. These elements also serve to explain the difference between perceptual and

imaginative states, as will become clear presently.

The phenomenon of perceptual constancy, though defined in various and

incompatible ways by different theorists, admits of illustration by uncontroversial

examples. There are two forms: synchronic constancy and diachronic constancy. An

example of the synchronic variety is this: when looking at a blue wall part of which

stands in shadow and part of which stands in sun, there is an obvious sense in which the

shadowed part of the wall looks darker than the sunlit part, and an obvious sense in which

the wall looks uniformly blue. An example of the diachronic variety is this: when comparing the way a blue wall looks on a bright, sunny day vs. on a dark, cloudy day, there is an obvious sense in which the wall appears not to have changed in color, and an obvious sense in which the color appears brighter on the sunny day and darker on the cloudy day.

Thus color experience presents objects that appear *both* uniformly *and* non-uniformly colored, or again, whose color appears *both* changing *and* unchanging over time. What explains the apparent contradictoriness of the contents of color experience in such cases? The resources at our disposal suggest a solution to this puzzle. Sticking to the synchronic case for now: during such episodes, color-features are presented to the subject in two different ways. The subject *instantiates* blueness of two distinct shades, and these properties are bound up with visual presentations as of different regions of the wall. At the same time, the subject *intends* blueness of a single shade. Color is presented *both* via an instantiated S-S property and as the intentional object of a P-I property.

There are a number of closely related cases in perceptual modes other than vision. (1) Volume-constancy: as a car approaches, the sound of its engine seems to grow in volume, in one sense, and seems to remain a constant volume, in another sense. Explanation: the S-S noise property bound up with the experience of the car heightens in volume (here *volume* is not a separable S-S property but one of the dimensions along with S-S noise properties vary), whereas a similarly bound P-I noise property remains unchanged over the same time interval. (2) Phone/phoneme-constancy: We can recognize sameness of syllable or word across great differences in pronunciation and vocal timbre: we can hear a British three-year-old and a Texan octogenarian as uttering the *same*

sentence. Explanation: the experience of differences between speakers is brought about by distinct S-S noise properties, while the experience of sameness is brought about by either identical P-I noise properties or else (what seems more likely to me) identical P-I *motor* properties. That is, it seems to me that when we recognize spoken-word types, we are recognizing them as types of *utterance,* rather than as types of *noise.*[18] (3) Letter/lexeme-constancy: We can similarly recognize sameness of letter and lexeme across differences of handwriting, font, etc. Here the explanation will appeal to fully determinate S-S shapes (the precise way that *this* 'A' appears) in connection with less determinate P-I shapes (the range of shape-appearances recognizable as presenting an 'A'.). Similar things can be said for recognition of (4) olfactory and (5) gustatory types— e.g., experiencing a particular odor as generally musty or a particular taste as generally bitter. Thus phenomena (1) – (5) all admit of the same type of explanation: each involves the simultaneous presentation of two features, one of which is instantiated and one of which is intended, both attributed to the same perceptually presented individual.

Why think that this is the right account of the phenomenon of perceptual constancy? A full defense is not possible here, as it would require a good deal of comparison with alternative accounts.[19] I will only venture what I take to be two explanatory pay-offs of my account. First, my account is able to sort out the ways that

---

[18] One reason to think so is the strong overlap between the neural substrates of speech-perception and those of speech-production (Liberman & Mattingly [1985]). Now, it is clear that neither P-I noise-types nor P-I motor-types cut finely enough to capture the perceptible differences between words: homophones (such as 'aloud' and 'allowed') are associated with identical motor- and noise-types. Speech-perception can thus involve perceptual categorization at a higher level than I am here discussing.

[19] Cohen's (2008) theory of color constancy bears important similarities to mine. On his theory, the experience of uniform or unchanging color in an episode of color-constancy is a matter of one's representing counterfactual properties of objects, i.e. how an object would appear under normal illumination conditions. While it is true that the color a surface appears to uniformly/unchangingly have is, typically, the color it would have under normal illumination conditions, I don't think that uniform color is presented *as* a counterfactual property of a surface. Perhaps perception presents objects as having counterfactual properties, but only in much more sophisticated forms of perceptual representation than perceptual constancy requires. Thus his proposal strikes as implausibly and unnecessarily complicated.

consciousness is simultaneously determinate and indeterminate. It is plausible, on the one hand, that there are no fundamental, indeterminate property-instances: the world is fundamentally made of determinate qualities instantiated by particulars.[20] If one holds that phenomenal properties are fundamental, as I do,[21] it follows that no instances of phenomenal properties are indeterminate, or at any rate if there are instances of indeterminate phenomenal properties, they are instantiated in virtue of the instantiation of determinate phenomenal properties. (For example, if I am instantiating generic painfulness, this can only be because I am instantiating some specific, determinate pain-quality.) But, on the other hand, perceptual constancies seem to involve features that are both indeterminate and, in some sense, phenomenal. This is obvious in the cases of olfaction and gustation I mentioned: when I recognize an odor as musty, the quality I thereby attribute to the experience is more general than the specific odor-quality I instantiate. But it is also the case in more canonical types of perceptual constancy. Consider the experience of casually glancing at a stretch of pavement. There is a color it will look to you—a grayish color, presumably. But if you attend to your experience, you will find that the stretch of pavement is presented as mottled and variegated in color, owing to its variation in texture and composition. Which shade of gray did the pavement initially look to you? I contend that there is an answer to this question, and it comes in the form of a *rough* shade. My theory has the virtue of resolving this puzzle of how determinables of phenomenal properties could be presented in consciousness: while instantiated (phenomenal) colors are determinate, intended colors, while no less phenomenal, can be less than fully determinate. (Perhaps they *must* be.)

---

[20] Notwithstanding attempts by Jessica Wilson (2013) and Elisabeth Barnes (2014) to give shape to ontic indeterminacy.
[21] More on this in chapter 6.

Such more-or-less determinate intendings of phenomenal qualities I'll call "P-I sensations." How do we come to instantiate them? One answer is that they are among the primitive P-I properties. Whether or not it's true that P-I sensations are *metaphysically* primitive (a question to which I return in section 6 below), they certainly need not be *developmentally* primitive. Which general types of sensations we can recognize as such seems to be a contingent matter. That it is a contingent matter at least partly explains the difference between the aesthetic connoisseur and the aesthetic novice: the connoisseur has developed the capacity to track more finely-grained visual, auditory, or gustatory sensations than has the novice. Perhaps, then, there is something to the old, concept-empiricist idea that we make "copies of impressions" which ground our ability to think about what we perceive. Not all P-I properties are such copies, of course, but many of them are. It would be misleading to think of them as *mere* "copies," precisely because they can be less determinate than the phenomenal episodes whence they derive: they must be *abstractions* out of such episodes. More precisely, P-I sensations present vaguely-bounded regions (perhaps *centered* regions) of a phenomenal quality-space.

In sum: the first explanatory payoff of my account of perceptual constancies is that P-I sensations help us make sense of the way in which indeterminate phenomenal qualities can be presented in experience, as is at least sometimes the case during episodes that exhibit perceptual constancy. Here's the second explanatory payoff: P-I sensations help us understand the difference between a perceptual episode in a particular sensory mode and an imaginative episode within that same mode. Suppose all of the S-S colors in a visual-perceptual state were replaced with P-I colors. I suggest that the result would be

a visual-imaginative state with precisely the same (narrow) content as the perceptual state.

Thinking about the relationship between perception and imagination in this way helps us to understand the core continuities and discontinuities between the two modes. Their core continuities consist in the fact that, in a somewhat elusive sense, imaginative episodes are facsimiles of perceptual episodes; images re-present the content of perceptual episodes in the very same perceptual mode. P-I sensations present mode-specific phenomenal qualities; hence this continuity.

Their core discontinuities are these: (1) Imagination can abstract away from perceptual episodes, presenting unsaturated particulars and properties.[22] (It's possible to imagine the car I saw parked in front of my house, and imagine it as yellow, but not imagine it as having a determinate shade of yellow.) This phenomenon is quite unmysterious, given that P-I sensations present determinables of phenomenal qualities. (2) Ordinary perception immediately and utterly compels our belief, whereas ordinary imagination does not compel our belief in the slightest. The difference between P-I sensations and the S-S properties they "copy" explains this. Consider the difference between felt pain and imagined pain: no matter how vividly we imagine pain, there is a type of *manifestness*, an unavoidable here-and-now actuality, that imagined pain lacks. A similar contrast exists between perceptual sensations and corresponding P-I sensations. The contents of perceptual experiences are anchored in the here-and-now in virtue of being bound to S-S properties. For example, the phenomenal redness I instantiate (say) is bound up with shape- and location-features in a P-I property cluster; hence I seem to be

---

[22] Compare McGinn (2005), ch. 1. McGinn lists nine sources of contrast between perception and imagination.

presented with a—vividly, manifestly, currently existent—red-colored particular. But if

corresponding mode-specific sensory features are *intended* rather than *instantiated,* no

such anchoring occurs—hence the discontinuity.[23]

P-I sensations thus amount to a derivative, yet narrow, level of phenomenal

intentionality. Before moving on to discuss wide contents, I need to say a brief word

about two types of perceptual constancy that are *not* to be explained in terms of the joint

instantiating and intending of a sensation: shape-constancy—e.g., a tilted coin looks

round, in one sense, and looks elliptical, in another sense—and size-constancy—e.g., an

approaching car appears to be growing larger, in one sense, and appears to remain the

same size, in another sense. Take the case of the tilted coin. In keeping with the pattern of

explanation above, I could say the following: a subject simultaneously instantiates S-S

ellipticality and P-I roundness. But whatever these two properties are, they do not bear

the same relationship to one another as do, say, S-S colors and P-I colors: P-I shapes are

not abstractable out of S-S shapes (if such there be). Here's why not: I believe that spatial

features, as presented in perception, are P-I primitives. For one thing, there is empirical

evidence that newborn animals perceive depth. For another, spatial features do not seem

---

[23] Two objections to the theory of the imagination I am developing here. Objection1: While it may be
economical to account for perceptual constancies and perceptual imagination with the same tools (viz., P-I
sensations), the view on offer has the following implausible consequence: there is no difference between
seeing the pavement as gray and imagining that the pavement is gray. Consequently, imagining that the
pavement is (contrary to fact) pink is no different from seeing it as pink.
Reply: Not so. Seeing the pavement as gray is a matter of the binding, via P-I attribution, of P-I grayness to
a property-cluster that visually presents the pavement. Imagining the pavement as pink involves a different
mode of semantic combination, something like *as-if-attribution.* Adding imaginative elements to one's
perceptual experience is a sophisticated task, involving whatever makes possible acts of pretense.
Objection2: Imaginative episodes can include constancy-phenomena. I can, for example, imagine a wall
that is uniformly white yet is partly in shade. But if imagination is built out of P-I sensations, then such
episodes would amount to attributions to the same object of contradictory properties.
Reply: I'm not sure that imaginative episodes can involve constancy-phenomena. If they can, I would be
inclined to bring Cohen's counterfactualist theory to bear here (see footnote 20): imagining a partly-in-
shade, uniformly-white wall involves intending whiteness and grayness, on the one hand, and intending
whiteness-under-normal-conditions, on the other. This, too, is a cognitively sophisticated task, not the sort
of task we should expect a theory of low-level intentionality to account for.

to be mode-specific: they can be bound to color-sensations in visual experience, noise-sensations in auditory experience, tactile-sensations in tactile experience, and motor and somatic sensations in proprioceptive experience. Thus, even if we could understand how spatial features could be abstracted from sensations alone—which I do not think we could—there would be no way to determine *which* sensory mode spatial features are abstracted out of. In short: the spatial features presented in perception are not subjective sensations. Rather, they are either types of (intended) relation between the subject and external objects (as is the case of egocentric spatial features) or else types of (intended) relation between points or regions of egocentric space (as is the case of allocentric spatial features, including shapes).

P-I shapes such as P-I roundness are not abstracted sensations, therefore, and so do not present vaguely-bounded, centered regions of a phenomenal quality-space. Rather, if they are not primitive, they are abstracted from other intended spatial features, and thus present more-or-less-sharply bounded region-types of egocentric space. It is properties such as these that ground the experience of constant shape in cases of shape-constancy, and the experience of constant size in cases of size-constancy. What makes for the experience of size- and shape- *difference*? The right answer appeals to *portions of the visual field*. The sense in which a tilted coin "looks elliptical" is that the presented coin occupies a roughly elliptical portion of the visual field. The sense in which an approaching car "looks larger" than it did a moment ago is that the presented car occupies a larger portion of the visual field than it did a moment ago.[24]

---

[24] There are oddities to this treatment of size- and shape-constancy. First, it dissociates size- and shape-constancy from other forms; second, it posits a "visual field," i.e. a unified structure of visual sensations (but not unified via any binding relation we have so far discussed); third, parts of the visual field seem to bear spatial relations to one another yet these spatial features are not reducible to *represented* spatial

4. Third Tier: Individuals.

So far I have exclusively discussed those features primitively presented in consciousness, either by instantiation (in the form of S-S properties) or by intention (in the form of P-I properties). In the last chapter I discussed how these features can be semantically combined in P-I property clusters—phenomenal structures comprising S-S and/or P-I properties bound (via P-I attribution) to P-I units. Now, while all the features presented to a subject in a P-I property cluster are grasped by the subject, these grasped features need not exhaust the intentional contents expressed by that cluster: these features can together serve as the descriptive mode of presentation *of a particular*. For example, if I instantiate a P-I property cluster that presents an object roughly the shape, size, and location of my brother, and that includes S-S colors my instantiating of which my brother is causally responsible for in the canonical way, then I see my brother (or, at any rate an entity momentarily coincident with my brother). Supposing there were individual essences and that my brother had one, I certainly do not grasp his essence. Thus my brother is part of the wide content of my perceptual state.

Now, my brother can be part of the intentional content of my perceptual state without my *recognizing* that he is; for all I've said so far, I do not yet see my brother *as* my brother. Indeed, for all I've said so far, I may not yet see my brother as a persisting object (in which case perhaps I do *not* see my brother, rather than a *stage* of my brother, or some such exoticum). So there is more that goes into representing a particular *as the*

---

features. I am inclined to take these oddities on board, in the absence of a phenomenologically adequate alternative. One benefit of the present proposal is that it explains the ambivalence some theorists have felt about size-constancy. Schwitzbegel (2011), for example, surmises that if tilted coins looked elliptical, then in some sense the world would have to look flat. I can respond: the sense in which coins can look elliptical does not involve the attribution of any geometric feature to them. The world does *not* look flat.

individual that it is.[25] To anticipate, what is required is that a subject construct a referential apparatus with respect to an individual, via the formation of phenomenal prototypes—a process which ultimately results in the subjects' *conception* of an individual. There are several key components in this process, including the experience of diachronic persistence, sorting, perceptual memory, and numerical identity. I'll discuss each of these components in turn.

First, a subject has to have a capacity for tracking persisting objects over time. In the first instance, this will mean that when a subject instantiates a P-I property cluster in a perceptual state over a time interval, the particular thereby presented will seem to endure through that time interval. I do not think this means that the subject needs explicitly to represent the particular *as* enduring. (Indeed I do not think the inclusion of such content could ground the experience of object-persistence, if that experience were not already present.) Such must be part of the nature of P-I units: not just that they particularize contents but that they present particulars as enduring, and hence guide the subject in making judgments about identity over time ("this frog is the same as the frog I was looking at a moment ago"), for subjects cognitively capable of making such judgments. Further, P-I units need to supply such an experience of the endurance of particulars across changes in the properties bound to them. Selfsame things are continually changing in their appearances, if only because they are presented as nearer than they were a moment ago.

---

[25] The matter is complicated; it depends on how the two constraints on a perceptual reference—accurate representation and causal ancestry—are weighted. If the causal constraint is heavily weighted, and if exotica such as brother-stages aren't good candidates for being causes of perceptual experiences, then perhaps I see my brother even if I am incapable of representing him as a persisting object. I am inclined to put more weight on the representational constraint and less on the causal constraint. My treatment of perceptual reference in this chapter leaves room for some disagreement here, however.

Second, experienced particulars need to be *sorted*: they need to be experienced as a certain sort of thing. That is, a subject's capacity to experience particulars as enduring leaves open the matter of *the conditions under which* a given particular is experienced as enduring. A thing's sort determines its apparent persistence conditions: a lump can persist through changes that the statue it composes cannot, so a thing seen as a lump will be seen as persisting through changes that a thing seen as a statue will not be seen as persisting through. A thing's sort also determines what sorts of changes are normal for it. Consider the difference between the behavior of a tree, of a balloon, and of a firefly: a thing thought of as a tree would be very difficult to recognize a moment a later, were it to behave in canonically balloon-like ways. Tracking a firefly as it moves through a field on a summer night would be nearly impossible if it were thought of as a very different kind of thing.

In short, recognizing numerical identity across time depends on taking things to be of certain sorts. Now, we have not yet discussed how cognizers such as ourselves come by sortals such as tree, balloon, and firefly (we'll return to this matter in the next section), as such contents are, plausibly, derivative. There must be *some* primitive sortals, however, to explain our most rudimentary sense of object-persistence, and we have already been introduced to some likely candidates. For example, according to Spelke's research on core cognition discussed above, there are basic ways we expect physical objects, qua physical objects, to behave, including "*cohesion* (objects move as connected and bounded wholes), *continuity* (objects move on connected, unobstructed paths), and *contact* (objects influence each others' motion when and only when they touch)."[26]

---

[26] Kinzer & Spelke (2007), p. 257.

Because we bring presented particulars under this sort, we are able to track them through changes in their appearances.[27]

What does sorting come to, metaphysically? One possibility is that attributing a sortal to a presented particular is a matter the inclusion of a P-I property whose intentional object is that sortal-property, bound to a subsumed P-I property cluster:[28] for example, experiencing a presented particular as a "Spelke object" would be a matter of P-I objecthood's being among the features attributed to it.[29] If this is the right picture, then sortals have to be explicitly represented. Another possibility is that a sortal need only be *implicitly* represented. I implicitly represent an experienced object as of a sort if I expect the object to behave consistently with things that of that sort, and hence I would react with surprise were it to behave inconsistently with things of this sort.[30] I am inclined to go this route. It is, after all, a ubiquitous phenomenon that we can react with surprise to unexpected experiences, without having been explicitly representing what was expected. (I reach out for and grasp my pen, thereupon reacting with surprise at its massive weight. Had I been I consciously, explicitly under-representing its weight? There is no reason to think so.)

---

[27] Perceptual constancies are an indication that a type of sorting has already taken place. To see a partially-shadowed surface as uniformly colored is to see its manifest color as non-definitive of what sort of thing it is.

[28] See chapter 4, section 4.

[29] This would need to be a special sort of binding, inasmuch as sorting attributes essential, or at any rate inductively deep, features to a thing. There are several ways this unique type of binding could be accounted for. There could be different variants of P-I attribution, corresponding to different degrees of modal strength of the attribution. *Being of a certain sort* could amount to a P-I relation that connects a property-cluster and a P-I sortal. Or the modal strength of the binding could remain implicit: it could be an especially resilient bond, such that a subject is disposed to accept un-bindings (or negatings) of the other features in the cluster before she is disposed to accept the un-binding (or negating) of the sortal.

[30] In chapter 3 I discussed Masrour's suggestion that "implicit anticipatory conditionals"—essentially, how a subject expects conscious appearances to unfold—can play a central role in a reductive account of phenomenal intentionality. While I reject Masrour's wholesale reductionism, I do think that implicit anticipatory conditionals ground many types of conscious contents. Probably such contents will be less than fully determinate.

A third step[31] toward representing individuals is being able to remember past experiences of them. I understand memories of perceived individuals as imaginative episodes of a unique kind. A subject S's conscious episode M counts as a perceptual memory of an individual O under the following circumstances: (a) it is caused in the right way by a past perceptual state P (or by multiple such states) that included a P-I property cluster that was in turned caused in the canonical way by O, (b) it includes a P-I property cluster that (roughly[32]) replicates the cluster in P, but with S-S properties swapped out for P-I sensations (such that it no longer presents O as part of manifest, current reality), and (c) it includes phenomenology as of the cluster's being familiar. In other words, I am remembering a perceived object if my imaginative episode feels like a memory of it and is in fact caused (in the right way) by my past perception of it. Thus there is both a phenomenological and a causal condition on perceptual memory just as there is both a phenomenological and a causal condition on perception. I leave unspecified what this causal condition on perceptual memory is, just as I leave unspecified what the causal condition on perception is, but I take it that it will be subject to the same sorts of constraints: it will have to be non-deviant, i.e. robust enough across counterfactual scenarios to deliver a sufficient degree of reliability. It is the phenomenological condition—the feeling of familiarity—that presses into referential service the causal condition, thus anchoring its reference in past causal interface with the world.[33]

---

[31] Though this step is typically involved in process of constructing a referential apparatus for an individual, I don't think it's required: it's possible to make the leap from a perceptual representation of a particular to an imaginative prototype of an individual.

[32] I can remember my having perceived O while misremembering some, but not all, of the features whereby I had perceptually picked out O.

[33] Another way to put the point is that without such anchoring, imaginative P-I clusters present the subject with the phenomenal equivalent of *indefinite descriptions*. Such anchoring is required for them to present the subject with the phenomenal equivalent of *definite descriptions*.

Once these elements—(1) the experience of diachronic persistence, (2) sorts, and (3) perceptual memory—are in place in consciousness, the subject is in a position to form a *phenomenal prototype* of an individual. A phenomenal prototype is an imaginative P-I property cluster that binds together those features of an individual that the subject uses to pick out that individual—including (explicitly or implicitly) the individual's sort, as well as the individual's prototypically identifying features. Though a prototype is not a memory of any particular perceptual encounter with an individual, it is abstracted from one or more perceptual memories, and thus it traces its causal ancestry through one or more perceptual experiences of an individual. Just as a perceptual memory differs from a mere imagining by involving the phenomenology of familiarity which serves to press into semantic service that memory's causal ancestry, so phenomenal prototypes involve a similar phenomenology of familiarity that has the same function, at one more causal remove. (It is not inappropriate to speak of phenomenal prototypes as a type of memory. One rightly answers the question "Do you remember so-and-so?" in the affirmative, if one can bring to mind a phenomenal prototype of that person.) Thus a phenomenal prototype is the phenomenal equivalent of a definite description, of the following form: the familiar [sort] such that [identifying features].[34]

For a phenomenal prototype to succeed in referring to an individual, its descriptive content plus causal ancestry need to be sufficient to uniquely pick out the individual. If I instantiate a phenomenal prototype that presents the content *that familiar physical object that is red and octagonal and has S-T-O-P written on it*, I will not thereby have succeeded in referring to a particular stop-sign, though a toddler who has only ever

---

[34] While the term 'prototype' is standard in cognitive science (for discussion see Laurence & Margolis (1999) p. 27ff), my use is non-standard inasmuch as I am treating prototypes as picking out individuals rather than picking out kinds. My usage is only very gently revisionary.

seen the stop-sign at the end of her street may very well be able to refer to it with just such a phenomenal prototype. This is not to say that for older animals that have seen a bit more of the world, fairly simply prototypes are never adequate. For example, it's plausible that my dog has all he needs to uniquely refer to me, if he can instantiate the phenomenal prototype that presents the content *that familiar animate object that smells thus-and-so* (even if he has sniffed scores of humans). While humans do not have such a power of smell, we do have the power to discriminate faces exceptionally well. It's plausible that I have all I need to refer uniquely to my brother, if I can instantiate the phenomenal prototype *that familiar person whose face looks thus-and-so*.

So, if a phenomenal prototype's descriptive content is sufficiently rich, or if the subjects' causal exposure to individuals is sufficiently poor, the sorts of narrow contents presented in a single prototype with pretty simply contents can mediate reference to that individual. But ordinarily the picture is much more complicated. For one thing, while we are probably hard-wired to home in on uniquely identifying facial features during very brief perceptual encounters, when it comes to non-human individuals, repeated perceptual encounter from different perspectives and/or awareness of inter-relations among remembered individuals is probably required for us to home in on uniquely identifying features. ("That familiar object which is ovoid and rough and gray" is not adequate to pick out the eighth pebble that caught my eye the first time I visited St. Columba's Bay. Nevertheless it is possible for me to refer to it—I just did.) For another, we generally have multiple ways of picking out the same human individual (voice, gait, dress, etc.). To begin to get these more complicated referential mechanisms up and running, a subject needs to be able to entertain two co-referential P-I property clusters at the same time—for

example, by simultaneously perceiving an object and remembering it—and for the two

clusters to be presented *as* co-referential. This could happen via the inclusion of a P-I

relation whose intentional object is numerical identity, and which links the two P-I

property clusters. *Naming* may have a crucial role to play here. "Common labels invite

comparison and abstraction," as Dedre Gentner's research demonstrates.[35] Sameness of

name strongly cues the experience of numerical identity.[36,37]

Under such circumstances, a subject is in a position to compare distinct

phenomenal modes of presentation of the same individual. This process of comparison

can shape prototypes in several ways. First, components from both modes of presentation

could be combined together to form a more robust mode of presentation. For example,

two experiences of the same face, one straight-on and one of its profile, could be used to

construct a prototype that includes a richer presentation of the three-dimensional shape of

the face than does either experience individually. Second, components of each could be

abstracted to form a more distilled mode of presentation. This could happen when a face

is experienced as bearded on one occasion and clean-shaven on another; the resulting

prototype will include neither beardedness nor clean-shavenness, but instead will include

the sorts of structural/relational features of the face common to both ways it appears.

Phenomenal prototypes can thus have a less determinate content than any particular

---

[35] Gentner (2010), p. 763. See also Waxman & Booth (2001) and Booth & Waxman (2003).

[36] Names, as linguistic types, are P-I sensations (as discussed in the previous section), i.e. abstract determinables of auditory- or motor-sensations. But of course they comprise a unique type of P-I sensation, a type that is experienced *as referring*, or anyway as *purporting* to refer. The signification-relation must therefore be presented in consciousness, serving to connect P-I properties that present names, on the one hand, with presentations of the things so named, on the other.

[37] Common labeling isn't the only psychological clue to numerical identity. (It had better not be, lest names had to exist before anyone could refer to the individuals named by them.) Certain perceptual features in common (e.g. sameness of scent) might raise the probability that two P-I property clusters are experienced as presenting the same individual. Or there might be contextual cues: a perceptually-present particular might seem to be the same as a remembered particular if its perceived appearance and location are just as one would expect of the remembered particular's to be, given its sort.

perceptual experience or memory, in two different ways: they can include a proper subset of the features presented in perceptions/memories of an individual, and they can present determinables of these features.

An additional possibility is that no *single* prototype emerges from a process of comparison of two modes of presentation of the same individual. Insofar as two modes of presentation make use of very different sorts of features of the individual—for example, how a thing looks from the top vs. the side, or two incompatible ways for it to behave— the comparison may issue in two distinct phenomenal prototypes as of the same individual. Consider, for example, my ability to think about *my house*. One prototype I can deploy in thinking about my house presents its visual appearance from the street: *that familiar structure shaped and colored thus-and-so.* Other prototypes involve its visual appearance from inside any of its rooms, or experiences of working in my study, of cooking in the kitchen, of a particular dinner party I once hosted, etc. If suitably related to one another, these numerous prototypes comprises my *referential apparatus* with respect to my house. Suitable relatedness includes at least my treating pairs of such prototypes as presenting numerically identical individuals and there being an associational link among my dispositions to instantiate these prototypes. More precisely: a subject's referential apparatus with respect to O is constituted by a set of dispositions to instantiate members of a set of phenomenal prototypes when (a) the subject is disposed to experience arbitrary pairs of such prototypes as presenting numerically identical individuals, (b) the instantiation of one such prototype occasions, or makes more readily available, the instantiation of others in the set, and (c) O is the best candidate satisfier of the conjunction of all the contents expressed by the members of the set.

The richer a subjects' knowledge of O, the more likely it is that her referential apparatus will involve a great many phenomenal prototypes. All of these prototypes are relevant to determining which object a subject is thinking of when she deploys one of them—despite the fact that the majority of them will remain dispositional most of the time. In cases of large referential apparatuses largely consisting of accurate prototypes, reference to O will be overdetermined. But a referential apparatus that is smaller or that contains inaccurate prototypes may fail to uniquely refer, or may be such that it is indeterminate whether it uniquely refers.

There is another, very different but crucially important, way that I can refer to my house that I have not yet discussed: I can think of it in terms of its *address*.[38] That is, I can think of it as *that familiar structure located* ____, where the blank is filled in either by reference to nearby landmarks, or by reference to its location in the broader (centered) world.[39] In employing relational/locational features of my house to identify it, I leverage reference off of a large chunk of my theory of the world. This theory will make reference to other individuals, and each of these individuals will require its own referential apparatus. Hence referential apparatuses for distinct individuals may be interdependent.

Once a referential apparatus is in place, I can pick out an individual in thought. I can then target the individual for further inquiry, proceeding to learn more about the individual. As I do so, a broader network of beliefs (i.e. dispositions to instantiate phenomenal structures that express conscious judgments) will form around the referential apparatus. This is my *conception* of the individual. My conception of the individual will

---

[38] For someone who lives in a subdivision consisting of dozens of identical-looking houses, this piece of prototypical knowledge will be an especially important component of her referential apparatus.
[39] Locations are themselves a type of individual. I can refer to locations insofar as I can consciously represent their spatial relation to other locations, and ultimately to my own location (current or remembered).

include my prototypical beliefs about it, but will also include other beliefs, running the *de re* modal gamut—i.e. beliefs about the individual's most transient features to its most essential. (I have learned, for example, that my house needs an exterior paint-job, on the one hand, and that it was built in the 1910s, on the other. While neither of these beliefs is particularly useful in fixing reference to my house, the latter involves the attribution of a feature more plausibly essential to my house than many of the features I actually use to fix reference to my house.) Over time, my conception of an individual may inform and revise the referential apparatus around which it is structured. Some of the prototypes that partly constitute my referential apparatus will drop out of the picture, some will be updated, and still others will be newly created.[40]

I have said that the whole referential apparatus is relevant to determining which individual a subject is thinking of, even though only a small part of the apparatus will be occurent at any one time: a single prototype can stand in the referential stead of the whole network. This is most obviously the case when an occurent prototype presents the individuals' *name*, e.g. *that familiar person named 'Cheryl'*. But there is a central type of mental reference to an individual in which none of the prototypes is, apparently, instantiated by the subject: perceptual recognition. When, for example, I recognize that the person walking toward me is my brother, there is a change in the content and the phenomenal feel of my overall experience, yet I need not instantiate an imaginative P-I

---

[40] For example, I don't remember the method whereby I was able to refer to George Clooney; it may have involved his name and the awareness that he was the dreamy star of a television show about nurses and doctors. But now that I have seen him perform and speak in many contexts (or rather have seen *videographic images* thereof), I am more inclined to bring to mind his vocal inflections and mannerisms— and these can serve to fix reference to the man even if I forget his name. They don't fix reference any *better* than 'The man named Clooney who stars in E.R.'; indeed, they seem to a worse job. (It's far more likely that there's a second man whose look and mannerisms are very close to Clooney's than that there's a second man whose name is 'Clooney' and starred in E.R.) On the other hand, they seem to get at features more essential to him than his name and his career-choices. To a certain extent, then, we seem to have a preference for essentialist prototypes of individuals. (Compare footnote 41 below.)

property cluster in addition to the perceptual P-I property cluster that presents my brother

to me, nor need I speak his name in inner speech. I suggest, rather, that what occurs in

such cases is that a placeholder that stands in for my entire conception of my brother is

bound to the P-I property cluster that presents him. This placeholder effectively renders

the perceptual P-I property cluster that presents him as a temporary member of my

referential apparatus with respect to him. The presence of this placeholder is marked by

three features: (a) a feeling of familiarity (i.e. that which serves to conscript causal

ancestry into referential service); (b) expectations with respect to the perceived

individual, in keeping with my beliefs about it—not just in accordance with its sort, but

also in accordance with the behavioral dispositions I am inclined to attribute to the

individual uniquely—and (c) the ready availability of conscious judgments from my

network of beliefs about the individual, especially in the form of other phenomenal

prototypes. There is, then, no phenomenology proprietary to seeing my brother as my

brother. But when I recognize my brother, a phenomenal shift does take place: the

phenomenology of familiarity is newly instantiated. And certain dispositions to think and

react in certain ways are activated.

5. Fourth tier: Kinds.

Being able to think of individuals is a matter of (1) remembering and comparing

perceptual episodes in which particulars are presented as enduring individuals of a certain

sort; and (2) on the basis of these capacities for perception, memory and comparison,

constructing prototype-based referential apparatuses. Being able to think of *kinds*

involves a process that is very similar. But it differs in two crucial ways. First, while

comparison of presented particulars plays a crucial role in generating and refining kind-prototypes (just as it does in generating and refining individual-prototypes), what gets compared are not modes of presentation of individuals that the subject treats as numerically identical, but rather modes of presentation of individuals that the subject treats as *type-identical*. For example, consider what it takes to think of cars as such. A subject has a perceptual encounter with a car, experiencing it under some sortal (object, artifact, mode of human transport, etc.). The subject may construct a prototype from just one such perceptual encounter; depending on the relative richness or austerity of the sortal the car is presented under, the prototype is apt to be more or less accurate in highlighting uniquely identifying features of cars. Multiple perceptual encounters will allow her to make comparisons, and hence to refine her prototype(s), in a process akin to hypothesis-testing. (Again, *naming* will substantially speed up the process.)

A second crucial difference is that the resulting prototypes will not serve to pick out individuals but rather to pick out kinds. How could this be? I can think of three possibilities. First, recall that prototypes are just a certain species of imaginative P-I property cluster, i.e., a bunch of features bound to a P-I unit. Such structures can pick out particulars because P-I units particularize the contents bound to them. But perhaps there is another type of P-I property cluster, one whose hub is not a P-I unit but rather a hub that preserves the generality of the features bound to it. We could call such hubs "P-I generics." When features are bound to them, the result does not even purport to pick out a particular, but rather presents the subject with a generic member of a kind. Whether or not the idea of P-I generics makes sense, I don't think this is how we use prototypes to pick out kinds, because I don't think I can make out a phenomenal difference between the

two types of property cluster. When I deploy a phenomenal prototype for the kind *horse* and when I make up a story about an imagined (particular) horse, I don't think there need be a difference between the P-I property clusters I use in each case. I think instead that we use representations of particulars—even when they don't seem to pick out any *specific* particular—to refer to kinds.

A second possibility is that there is no intrinsic difference between prototypes used to pick out individuals and prototypes used to pick out kinds. Instead, the distinction between individual- and kind- reference lies in the level of specificity of the content in a referential apparatus. The thought would be that referential apparatus pick out whatever individual *or set of individuals* best fits its descriptive content. Picking out a kind is just a matter of deploying descriptive content general enough that many individuals (viz., those that delimit the kind) satisfy it. But I don't think this is adequate, either. Suppose I encounter a flower of a type I don't recognize. I can form an individual-prototype, expressive of the content *that familiar flower with features such-and-such*. I can then use this prototype to pick out the *kind* of flower. The referential apparatuses are identical, but one picks out an individual and another picks out a kind.

So I think there must be a primitive, intrinsic difference between prototypes used for purposes of kind-reference and prototypes used for purposes of individual-reference. The difference could amount to an additional phenomenal element in individual-prototypes or in kind-prototypes, or in both; since we have discussed prototypes in their individual-referential function and have not found them lacking in anything, it is natural to look for the additional feature in kind-prototypes. If individual-prototypes express the content *that familiar [sort] such that [identifying features]*, kind-prototypes express the

220

content *that familiar kind of [sort] such that [identifying features]*. There must be, in other words, an element in consciousness that presents a particularized content as a token of a type. When this element affixes to a prototype, it de-particularizes the prototype, or rather, lets it stand in the stead of the category of items that match it qualitatively.

So, the referential apparatuses for kinds differ from referential apparatuses for individuals, in that (1) prototypes are refined on the basis of comparing modes of presentation of type-identical rather than of numerically identical particulars, and (2) such prototypes are used in a primitively kind-representing capacity. Otherwise the two types of referential apparatus are alike. Consider the process of learning about salt. A first encounter might be with a variety of tastes that share a quality in common, from which a prototype emerges that expresses the content: *that kind of stuff that tastes thus-and-so*. Visual experiences with salt might result in a distinct prototype or prototypes, depending on how much variation there has been in her visual experience of salt—depending on how much the salt crystals of her acquaintance have varied in size, say. Once this referential apparatus is in place (assisted by learning the word 'salt'), the subject is able to think of salt—though it may be indeterminate just which extension the referential apparatus picks out, as some of the prototypes may pick out extensions that do not perfectly overlap with one another. The subject can then target the kind *salt* for further inquiry, proceeding to learn more about it, including its chemical makeup. The broader network of beliefs that form around the referential apparatus constitutes the subject's conception of salt. As with one's conception of an individual, one's conception of a kind can inform and revise one's referential apparatus. Some of the prototypes that partly compose my referential apparatus will drop out of the picture. I could learn later that my

initial encounter with saltiness was actually caused by a different chemical—MSG, say—

than the one my referential apparatus came to settle around. Perceptual prototypes might

get replaced with theoretically-informed prototypes which employ essential rather than

accidental features of the kind.[41]

My account of how we use phenomenal prototypes to pick out individuals and

kinds has so far assumed that the starting point is perceptual encounter with a kind. But

of course we can form conceptions of individuals and kinds that we have never

perceived. One very common example of this is our ability to form conceptions of

individuals and kinds we have only seen *pictures* of. We can also form conceptions of

individuals and kinds we have only had described to us, and we can even form

conceptions of kinds we could not possibly perceive (quasar, electron). The causal

relations required to anchor reference for prototypes of kinds we *have* perceived is

complicated enough: a prototype of a kind K has to have been caused in the right way by

a perceptual state (or by memories that were caused in the right way by a perceptual

state) that was caused in the right way by an instance of K. But when what is perceived is

not an instance of kind K but rather a *representation* of kind K (be it an image or an

utterance or an inscription), then this representation has to itself have been caused in the

right way by instances of kind K. Such *deferential anchoring* of the reference of our

prototypes is, I suspect, as common as it is complicated.

Once one has formed a conception of a kind, that conception can seed the forming

of more abstract categories. For example, prototypes from one's duck-conception, one's

tree-conception, one's human-conception, and so forth could be type-identified in a new

---

[41] See e.g. Medin & Ortony (1989). We seem to have an innate preference for essentialist categorization, i.e. for constructing prototypes that track deep, hidden structures of a kind rather than mere surface ones.

referential apparatus that picks out the kind *organism*. Again, such abstract categories can serve as the sortals that form the backbone of prototypes for less abstract categories, such as the kind *fungus*. Thus, previously acquired conceptions serve as a scaffold for the acquisition of new conceptions, in an upwards direction (of increasing abstractness) and a downwards direction (of decreasing abstractness). And new sortal-categories allow the subject to more readily form conceptions of individuals. (I could perhaps form a referential apparatus that uniquely picks out my car without ever being able to categorize it as a car, but it would be much easier to do with a car-sortal at hand.)

My account of perceptual recognition from the previous section carries over, mutatis mutandis, to perceptual categorization. When I perceive a duck *as a duck*, a placeholder-property that stands in for my whole duck-conception is bound to the P-I property cluster that presents the duck. This placeholder effectively renders the property cluster as a temporary member of my referential apparatus for the kind *duck*, and its presence is marked by (a) a feeling of familiarity, (b) expectations with respect to the categorized individual, in keeping with my beliefs about ducks, and (c) the ready availability of conscious judgments from my network of beliefs about ducks. There is no proprietary phenomenology to perceiving something as a duck, though there is a concomitant feeling of familiarity and an activation of dispositions to think and react in certain ways.

I have spoken of conceptions; what of *concepts*? The term 'concept' is used in different ways by different theorists, so it isn't entirely clear what features an entity has to have to count as one. Nevertheless, I do think that if a subject has constructed a referential apparatus that picks out the kind *duck*, it is reasonable to say that the subject

"possesses the concept DUCK". What I have said about referential apparatuses is consistent with much of the conventional wisdom regarding concepts among psychologists and philosophers. Here are a few of the features often associated with concepts for which I have accounted:

(1) Prototypicality: For most kind-concepts, subjects judge that some members of a kind better exemplify the kind than others do.[42]

My explanation: reference to a kind is fixed by whatever is the best match with a plurality of prototypes. Inasmuch as prototypes capture the ways that members of the kind are most readily recognized, prototypes render certain identifying features more salient than others, and hence render members of the kind that exemplify such features more salient than members of the kind that do not.

(2) Non-definability: For most kind-concepts, subjects cannot provide adequate necessary and sufficient conditions for kind-membership.[43]

My explanation: reference to a kind is fixed by whatever is the best match with a plurality of prototypes, not by recourse to necessary and sufficient conditions.

(3) Theory-like structure: Many kind-concepts are constituted by complex inferential relationships and can evolve over time, as scientific theories can.[44]

My explanation: A network of prototypes is involved in fixing reference to a kind, and the boundary between the reference-fixing elements of a conception, on the one hand, and the non-reference fixing elements, on the other, is porous and dynamic.

---

[42] Rosch (1978), Lakoff (1987).
[43] Laurence & Margolis (1999), pp. 14-16.
[44] Gopnik (1996), Carey (1999).

(4) Linguistic reciprocity: language influences conceptualization but does not wholly determine it.[45]

My explanation: we can treat linguistic items as meaningful because we have intentional capacities that are psychologically prior to our linguistic capacities. But linguistic items help us construct conceptions, by (a) indicating type-identity across presented particulars (for purposes of comparison and subsequent prototype-formation), and (b) facilitating the deferential anchoring of reference.

(5) Semantic combinability: concepts are systematically recombinable with each other, according rules of generative grammar.[46]

My explanation: Prototypes can combine with other P-I properties to form structures that express propositional contents (via the mechanisms discussed in the previous chapter). Because a single prototype can stand in for an entire referential apparatus, the propositional content thereby expressed contains, as a constituent, the kind denoted by the entire referential apparatus.

(6) Shareability: it is common for a subject at two different times, or for two different subjects, to have the same the concept. [47]

My explanation: Though a subject's conception of a kind is not likely to be shared in every detail with other subjects at a time or with herself over time, if it is robust enough (if it affords a subject with an ability to think about a kind, to discriminate members of that kind from non-members, and to carry out these cognitive tasks in more than one way, etc.), it counts as a realizer of a *type* of conceptual ability that can be multiply realized. More or less finely-grained types of conceptual ability

---

[45] Spelke & Tsivkin (2001), Carey (2009)
[46] Fodor & Pylyshyn (1988), Fodor (1999).
[47] Fodor (1999).

can be shared. Inasmuch as humans have prototype-forming mechanisms in common, what is shared across subjects might be quite fine-grained indeed.

Social deference in linguistic use helps shore up similarity of conception as well. Thus my account has the virtue of ecumenically accommodating many of the desiderata used to evaluate theories of concepts.[48]

6. Fifth tier? New primitives.

I take it that the foregoing is a phenomenologically plausible, empirically adequate theory of large portions of our intentional capacities, or at least a sketch of the general shape of such a theory, that is consistent with the Phenomenal Grounding Thesis. But there are reasons to suppose that it is incomplete, because too bottom-up. Several types of intentional phenomena indicate that *new P-I primitives* emerge along the way. These phenomen include (1) the grasp of highly abstract sortals, (2) the grasp of derivative determinables, (3) purely cognitive phenomenology, and (4) leaps in conceptual development.

Highly abstract sortals. In the previous section I described how multiple kind-conceptions can be serve as inputs to the construction of a new, more abstract kind-conception: for example, conceptions of ducks, of flowers, of trees, and so forth can be used to form a conception of the kind *organism*. One of a subject's duck-prototypes can thereby also serve as an *organism*-prototype. (Which kind it stands in for will depend on

---

[48] One final hallmark of concepts that I have not yet accounted for: as I mentioned briefly in section 2, it is increasingly common for cognitive linguists to think of *metaphor* as an ineliminable aspect of our conceptual repertoire. For example, Lakoff & Johnson (1980) maintain that our concept of a scientific theory is at least partially a metaphorical extension of our concept of a building. I am not sure what more would need to be added to my picture in order to account for the phenomenon of so-called "conceptual metaphor": a distinctly metaphorical type of semantic relationship between referential apparatus and referent, perhaps? I hope to explore the matter in future work.

which other prototypes the subject is disposed to treat as presenting particulars that are type-identical with it.) But some sortals are so abstract that practically anything (or any things) can serve as a prototype for it. *Object, cause* and *motion* are paradigm cases of such highly abstract sortals, and plausibly, they are innate primitives (presented either explicitly or implicitly). But it isn't obvious that all such highly abstract sortals are innate primitives, e.g. *noun*, *verb*, *consciousness*, *thought*, *belief, desire*, *rule*, *word*, *sentence*, *kind, number*, *identity*, *similarity, property*, *relation*. It is not implausible that some of these sortals are such that (a) they are not innate but acquired and (b) their nature is directly grasped in consciousness (vs. grasped indirectly, via a network of prototypes).

Derivative determinables. My explanation (in section 3 above) of perceptual constancies and imaginative episodes involves P-I sensations, i.e. *intendings* of features initially *instantiated* such as phenomenal colors, sounds, tastes, and so forth. While instantiated features are always fully determinate, P-I sensations can present vaguely-bounded regions of a phenomenal quality-space. But there is a puzzle about how these new P-I properties are formed, because subjects do not actually instantiate phenomenal quality-spaces. Rather, their experiences present various point-sized regions in the space. P-I sensations are not intendings of mere disjunctions of qualities, but rather intendings of genuine determinable qualities.[49]

There is a similar puzzle related to derivative geometrical features. Unlike P-I sensations, P-I shapes, distances, and so on are not derived from S-S properties. Instead,

---

[49] Laurence & Margolis (2012) propose a mechanism for how representations of determinate white can be leveraged to form representations of a less determinate color. I don't think their proposal helps solve the puzzle under current discussion, however. For one think, they help themselves to the notion of a quality-space, whose status as an input into the abstraction-process I am calling into question. Second, they are interested in the forging of a new tracking-link between a representation and that which it represents, where as I am interested in the forging of a new P-I property.

geometrical features are primitively presented. Not *all* are primitively presented, of course—we have to learn to recognize the shape of Maryland, for example. And we *are* presented with geometrical space (or at any rate with whatever portion of egocentric space in which our perceptual experiences present objects), in a way that we are not presented with phenomenal quality-spaces; we can "carve out" regions of geometrical space in consciousness, in a way not available to us when it comes to phenomenal quality-spaces. Still, the abstraction of spatial features raises a puzzle of its own, at least when it comes to fuzzy regions. For example, I can have a visual experience of a rectangularly-shaped book, and thereafter be able to imagine that book's shape and size, without imagining its *precise* shape and size. This doesn't seem to be a mere selecting and disjoining of what's already given, but the generation of a new primitive: a fuzzily-bounded region of egocentric space.[50]

Purely cognitive phenomenology. I have said much in this chapter about perceptual and imaginative states but not much about cognitive states. I now fill in this lacuna. What is unique to perceptual states is that they involve the binding of S-S properties to P-I properties. (S-S properties thus anchor such states' presented contents in manifest, current reality.) What is unique to imaginative states is that they include P-I sensations. Cognitive states amount to the remainder of intentional states, i.e. those wherein sensations are neither instantiated nor intended. Contents that amount to the intentional objects of primitive P-I properties can be included in cognitive states, of course. But individuals and kinds not primitively grasped in consciousness can also

---

[50] Gentner (2010) says that one product of a subjects' comparing objects or scenarios for sameness of structure is that the subject "rerepresents" the common structure as the distinct structure-type that it is. I am not sure how to map Gentner's language onto my own, but on one plausible reading, Gentner's abstracted structure-types are now grasped as new primitives.

figure in cognitive states. The very same placeholders whose presence in perceptual states makes for perceptual recognition and categorization can be instantiated outside of perceptual states. Cognitive phenomenology thus comprises P-I properties (excluding P-I sensations), placeholders for conceptions, plus additional semantic machinery such as P-I units, P-I attribution, credence-halos, inference-links, phenomenal quantifiers/connectives/operators, and the phenomenology of familiarity. Mine is thus a relatively scanty picture of cognitive phenomenology: placeholders are not discriminable from one another on the basis of their intrinsic natures, and so the contribution they make to fixing the content of a cognitive state is a matter of the imaginative states they *dispose* the subject to instantiate. Now, there are introspective phenomena that such scantiness nicely explains. For example, it is frequently the case that I have a conscious thought that I later struggle to put into words; the struggle can even reveal to me that there was no coherent thought to be had. My view allows for such poor grasp of one's own thoughts, while still endorsing a broadly Cartesian picture of introspection.[51]

But is my picture *too* scanty? An important source of motivation for the Phenomenal Grounding Thesis comes in the form of arguments for the existence of cognitive phenomenology (i.e. what it's like to entertain an intentional content, where this doesn't involve sensory or imagistic phenomenology).[52] It's worth asking at this point whether the theory I have been developing in the present chapter can supply a phenomenology of cognition rich enough to get these arguments going.

---

[51] See Siewert (1998), p. p. 276ff and Pitt (2011) for discussions of such phenomena and how advocates of richer cognitive phenomenology can accommodate them.

[52] Most of these arguments we explored in chapter 2. There I lumped together arguments that support the supervenience of intentional content on the phenomenology of imaginative states, with arguments that support the supervenience of intentionality by the phenomenology of cognitive states. I did so because the most heavily contested boundary in the phenomenal intentionality literature is the boundary between the perceptual and the non-perceptual. On my view, the boundaries between all three types of state are rather porous and fuzzy.

Phenomenological data from which these arguments proceed include (a) the phenomenal difference between not understanding vs. understanding a language, between understanding an ambiguous sentence in two different ways, and between the before and after of having a sudden, wordless and imageless realization; (b) our ability to introspectively discriminate thoughts, and (c) the manifest interestingness of thinking.[53] It is not obvious to me that my theory can't explain these data, but I think it's a live possibility that it can't. (My picture fairs most poorly with respect to linguistic disambiguation and manifest interestingness, I judge.) If it does not, new primitives may be in order.

Leaps in conceptual development. Susan Carey (2009) argues, on the basis of experimental evidence in developmental psychology, that the conceptual resources of adult humans are not completely continuous with the conceptual resources of children; at various points in the developmental process, children have to "bootstrap" their way into new primitive concepts. Carey's paradigm cases for such conceptual leaps are the concepts *number* and *material object*. Drawing an analogy with conceptual leaps that have taken place in the history of science, Carey claims that adult concepts of number and material object are "incommensurable" with precursor concepts with which children are innately endowed. Carey's argument for this incommensurability claim involves two empirical observations. First, children can manifest a partial facility with mathematical and folk-physical terms, while also failing at basic tasks in (what appear us to be) very weird ways. For example, a child can be capable of following instructions that involve

---

[53] "If there were no cognitive phenomenology, *life would be boring*—more boring than it actually is, at least. In particular, it would be quite irrational to engage in philosophical reflection if none of it 'showed up' in consciousness (surely philosophy is not as interesting as it is purely in virtue of affording the imagery it does." (Kriegel [2015], p. 40.)

discriminating groups of one, two, or three things, *and know how to count*, yet be totally unable to tell larger groupings apart. Or again, a child can have some sense of the relative weight of material objects, while expecting objects to weigh less when tipped on their side or when flattened. Second, getting from partial facility to full facility with mathematical and folk-physical terms is a very difficult, protracted task for children, spanning months or years. This is what we would expect, Carey argues, from a non-linear developmental process. Now, Carey's claims are not uncontroversial; some scholars think that linear developmental processes might yet explain the domains of interest to Carey.[54] But other scholars think that Carey has *underestimated* the amount of discontinuity between the child's conceptual endowments and the adult's.[55]

Now, it isn't immediately clear what the relationship is between Carey's primitive concepts and my primitive P-I properties. But if Carey is right that the acquisition of new concepts requires new raw materials and not merely the new combination and deployment of old raw materials, then the implication seems to be that new P-I properties will be involved in the construction of Carey's new primitives concepts. So there is some empirical support, in addition to the philosophical-cum-introspective support adduced in the previous few paragraphs, for the idea that the set of primitive P-I properties discussed in section 2 above is not adequate to construct all the intentional contents we adult humans can entertain.

---

[54] E.g. Gelman (2011), Gopnik (2011), and Spelke (2011).
[55] Cf. Kiss (2011), who thinks that adult concepts of mind are incommensurable with the child's concepts of mind. Elsewhere, Carey herself hints that the forging of new primitives might be quite widespread in human cognition: "take the features that determine the prototype structure of bird concepts (flies, lays eggs, has wings, nests in trees, has a beak, sings, etc.). Subjects provide distinctive values for such features when asked to list the features of birds, and overlap in terms of these same features predicts prototypicality within the category bird. That is, this feature space definitely underlies adult prototypicality structure. Yet these features are not innate primitives; many are no less abstract and no less theory-laden than the concept bird itself. One of the goals of TOOC [her 2009 book] is to characterize a learning process through which new primitives come into being (Carey [2011], p. 114.).

It is at any rate worth asking how new primitive P-I properties could get into the picture, were there any. Carey proposes a mechanism she calls "Quinean bootstrapping." It involves three elements: the constructing of a model of the target domain out of existing conceptual resources, the learning of a set of placeholder terms, and a drawing of an analogy between the relations among the components of the model, on the one hand, and the terms in the set, on the other. For example, children can model quantities by thinking of groupings of one, two, three, or four individuals, and they can learn to recite the numerals in order. At some point, a child notices that the numerals '1', '2', '3' and '4' are applied to perceived groupings of things that can be put in a one-to-one match to the elements of the groupings in their mental model.[56] The output of the process is the representation of natural numbers as such, i.e. as indefinitely generable via the successor-function.

How illuminating is Carey's proposal? It depends on her ambitions. If she is interested in explaining how children pick up the rules for a new language game (how/when one ought to utter number-terms, say), her proposal might be quite successful. But her ambitions are constrained by a less deflated notion of concept-possession, as I read her. Another interpretation might be that she is explaining the metaphorical grounds for our number-concepts. That is, perhaps our grasp of the successor-function is irreducibly metaphorical: to be the natural-number successor of 1 is to stand in a relation to the number 1 akin to the relation between a group of two things and a single thing. Again, I think this would be interesting, but not what Carey has in mind.

---

[56] Carey moves quickly between person-level and sub-personal-level talk; I'm not sure whether she thinks that any of this process involves conscious awareness or agency on the part of the child.

If, on the other hand, Carey is interested in explaining the mechanism whereby we come to *grasp* the numbers and/or the successor-function, I don't think her proposal goes very far. She has not provided a recipe for generating a grasp of a new idea as the output of a process whose inputs are old ideas, the way a chemist might describe how one generates salt as the output of a chemical process whose inputs are sodium and chorine. Now, Carey is explicit that her proposal does not amount to a *deductive* process; it involves an "inductive leap".[57] But the trouble is not that the process of generating new primitives is non-deductive. The trouble is that the output of the process is something genuinely new, an intentional content not expressible in the vocabulary of the inputs. And this suggests that there really can't be a mechanism for the generation of new primitives, at least not if this is a *rational* or *intentional* mechanism. There could, of course, be a *causal* mechanism. And that is perhaps what Carey has provided: the causal conditions under which a subject comes to instantiate new P-I properties she was not in a position previously to instantiate. Carey's talk of bootstrapping as an "inductive" process obscures the fact that the transition at issue is not a matter of shifting which hypothesis one endorses, but a matter of acquiring the intentional capacities whereby a hypothesis can be entertained at all. It is akin to the shift that takes place as a data-points get filled in on a graph until one suddenly sees not the dots but the line-of-best-fit through the dots—but without one's having had the notion of a line before. The difference between a theory of the *construction* of new derivative contents out of old primitive contents, on the one hand, and a theory of the *causal production* of new primitive contents by old primitive contents, on the other, is closely analogous to the difference between broadly reductive and broadly non-reductive theories of mind and body—to which I now turn.

---

[57] Carey (2009) p. 477.

6. The Emergence of Phenomenal Intentionality

Introduction.

A bit of review is in order. This dissertation has been motivated by the following
questions: What is mental content? And how can our mental states have it? I regimented
these semi-technical, semi-intuitive questions into the fully technical question in
metaphysics: What is the nature of intentional properties? My answer came in the form of
the Phenomenal Grounding Thesis, according to which all intentional properties are
partly grounded in phenomenal properties. In order to sell my answer as the *correct*
answer, I have had to provide (a) arguments in its defense, and (b) a picture of how it
could serve explain to the intentional properties we actually instantiate.

My *defense* of PGT occurred in three steps. First, I argued (in chapter 1) that non-
phenomenalist theories of intentional properties, i.e., those that appeal to
functional/dispositional facts alone, cannot account for the way that intentional properties
guide a subject's discriminatory judgments about sameness and difference of contents
across mental states. Second, I favorably reviewed a number of arguments (in chapter 2)
that point to a close modal connection between consciousness and content. Third, I
argued (in chapter 3) that there are no incontrovertible counterexamples to PGT.

My *explanation* of PGT involved positing primitive P-I properties (in chapter 3),
followed by an extended discussion (spanning chapters 4 and 5) of how a sparse set of
these properties combine, in perception, imagination and cognition, to form structures
that present propositional/descriptive contents. Some such structures serve as narrow
modes of presentation of wide contents.

There is an important worry that will come up for many readers at this point. In order to make sense of content, I have appealed to consciousness. How helpful is the appeal? A simple dilemma for my view purports to show that the answer is: *not very*. Here is the idea: either the problem of locating consciousness within a unified, empirically-informed metaphysics is a hard problem[1], or it isn't. If it isn't, then consciousness "reduces" (in at least one of that term's many senses) to the non-conscious goings-on. In that case, even if PGT is true, it is at best a way station; a fully satisfying metaphysics of intentionality will leave out consciousness-talk and instead appeal to whatever it is that consciousness reduces to. On the other hand, if consciousness is as hard to reduce as it is often taken to be, then I may not have made much progress, either: I have tried to explain one mysterious phenomenon in terms of a still more mysterious phenomenon.

In response, my strategy is to embrace the second horn, but to try to mitigate the mystery a bit. That is, I think that the problem of consciousness is hard in that it cannot be solved without adding to our ontology. But it is nevertheless possible to locate consciousness within a unified, empirically-informed metaphysics. I'll say briefly, in the remainder of this section, why I think that the problem of consciousness is hard. In the rest of the chapter I'll discuss three choice-points for a non-reductionist theory of consciousness. Few of the considerations I bring to bear are without precedent in the literature on the mind-body problem. Nevertheless, we will find along the way that the existence of P-I properties sheds new light on some of these considerations.

Phenomenal properties are, I believe, fundamental properties. A minimal list of the properties that jointly determine what our world is like will include them. Such is the

---

[1] 'Hard problem' as a technical phrase comes from Chalmers' (1995).

positive characterization of my view. It is more frequently identified by the way it is negatively characterized, i.e. as a form of phenomenal *anti-physicalism*. The negative characterization is only as good as our understanding of the category of a physical property. As a metaphysical category, I think we have a very poor understanding of it. But we can understand it as a methodological category, viz., as the category of properties that enjoy good standing in contemporary physics.[2,3] Here, then, is a composite (positive and negative) characterization of my view: phenomenal properties are fundamental properties, not identical to, composed by or realized in any other fundamental properties (including any of those that enjoy good standing in contemporary physics).

A host of arguments for phenomenal anti-physicalism have been proposed and vigorously debated over the last thirty years. They fall into two abstract argument-types, which I will call "Cartesian" arguments and "Leibnizian" arguments. Cartesian arguments point to epistemic gaps between applications of physical concepts and phenomenal concepts. The most discussed examples include the Conceivability Argument[4], The Knowledge Argument[5], and The Explanatory Gap Argument[6]; there are others.[7] Cartesian arguments begin with a premise asserting the existence of an epistemic gap between physicality and phenomenality. (In the case of The Conceivability Argument, the gap amounts to *conceivable separability*; in the case of The Knowledge Argument, *non-derivability*; and in the case of The Explanatory Gap Argument, *non-*

---

[2] Why not in our *completed* physics? Because we have no idea which properties will be invoked in our completed physics.

[3] A lingering ambiguity has to do with whether the properties invoked by contemporary physicists should be understood as role-properties or realizer-properties. The ambiguity affects whether panpsychism is a form of physicalism or a form of anti-physicalism; see footnote 15.

[4] Often called "The Zombie Argument," pressed by Chalmers (1996).

[5] Jackson (1982).

[6] Levine (1983).

[7] e.g. The Modal Argument (Kripke [1980]), The Property Dualism Argument (White [2010], Horgan & Tienson [2001], Martine Nida-Rumelin [2007]), and The Argument from Revelation (Stoljar [2009])

explainability.) A second premise asserts that the existence of the epistemic gap (between the applications of concepts) entails the existence of an ontic gap, between properties. A third premise asserts that the existence of the relevant ontic gap is inconsistent with physicalism.

Leibnizian arguments point to ontic gaps between physical and phenomenal properties. Examples of such arguments include The Structure and Dynamics Argument[8] and The Unity of Consciousness Argument.[9] Leibnizian arguments begin with a premise asserting that phenomenal properties can have some feature that physical properties necessarily lack. (In the case of The Structure and Dynamics Argument, the relevant feature is *possessing a non-dispositional nature*. In the case of The Unity of Consciousness Argument, the relevant feature is *being phenomenally unified*[10]). A second premise is a version of Leibniz's law: if two items fail to share all of their features, then those items are non-identical. A third premise asserts that non-identity of the relevant properties is inconsistent with physicalism.[11]

While these arguments have been far from universally persuasive, they have generated an enormous literature (mostly in response to the Cartesian arguments), suggesting that at if they are unsound they are *interestingly* and *non-obviously* unsound. Among those physicalists who take phenomenology seriously, something of a consensus has emerged regarding the best way to resist Cartesian arguments. First developed by Loar (1990) and dubbed 'The Phenomenal Concept Strategy' (or PCS) by Daniel Stoljar

---

[8] Chalmers (2003).
[9] Hasker (1999).
[10] LaRock (2007) develops a related argument from conscious unity, where the relevant unity is not that of whole phenomenal states but rather the unity of particular gestalts, e.g. the experience of a red square at a location.
[11] I call Cartesian arguments 'Cartesian' because of their affinities with Descartes' arguments. I call Leibnizian arguments 'Leibnizian' because they deploy Leibniz's law. (Leibniz's famous Mill Argument against the mechanical theory of mind is probably best understood as Cartesian.)

(2005), the general idea is that the epistemic gap between physical and phenomenal concepts can be explained in terms of the differences between the two types of concept. Elsewhere I have argued at length against this proposal.[12] There is no space here to enter fruitfully into the debate; for present purposes I will simply treat anti-physicalism as the winner. I offer my proposals in the present chapter conditionally on the correctness of that assessment.

P-I properties can probably be used to make the anti-physicalist case stronger. They can be used to construct both a novel Cartesian argument and a novel Leibnizian argument. The novel Cartesian argument exploits the fact that there is an epistemic gap between any candidate physical/functional property, on the one hand, and any P-I property, on the other. That is, it is hard to imagine a physical or functional description of the brain (say) from which it a priori follows that content is being presented to a subject. If the existence of such an epistemic gap entails the existence of an ontic gap, anti-physicalism about P-I properties presumably follows.

The novel Leibnizian argument exploits an apparent difference between P-I properties and any candidate physical/functional property with which P-I properties might be identified. P-I properties ground a form of intentionality that is *intrinsic* to conscious mental states. But, as I argued in chapter 1, naturalistic-intentional properties—those physical/functional properties that come closest to being extensionally equivalent to intentional properties—lack the feature of being intrinsic to conscious mental states (even granting a functionalist analysis of consciousness generally). If this difference in 2nd-order property entails a difference in 1st-order property, anti-physicalism about P-I properties presumably follows.

---

[12] Woodward (ms).

2. First Choice Point: Panpsychism vs. Emergentism

If phenomenal properties are not physical properties, where do phenomenal properties come from? How are they related to rest of concrete actuality? A first choice-point relates to whether phenomenal properties are instantiated by the most basic building-blocks of reality, or whether they *emerge* from reality once those building-blocks are suitably arranged. A more ontologically conservative answer to this question is *panpsychism*; a more ontologically profligate answer is *emergentism*.

For my purposes, Panpsychism is the view that the ultimate physical constituents of reality (the "UPCs") instantiate phenomenal properties, and that the phenomenal states of macroscopic entities (such as ourselves) are constituted by these properties.[13] Further, the constitution-relation that holds between macro-level phenomenal states and the phenomenal properties of the UPCs is of a piece with the constitution-relation that holds between macro-properties and micro-properties generally: the constituted properties are automatically generated, with metaphysical necessity, whenever the constituting properties are suitably arranged. Chalmers (ms) calls this view "constitutive panpsychism"—in contrast with "emergent panpsychism," according to which the explanatory relation between micro- and macro-properties is not as modally strong. (I will briefly touch on emergent panpsychism below.)

The main virtue of panpsychism is its elegance. Just as the fluidity of water is constituted by the bonding-properties of the atoms in its constituent molecules, and just as the volume of a cheer at a sporting event is constituted by the volume of the cheer of

---

[13] In assuming that there are UPCs, I am assuming that physics has to posit thing-like fundaments—particles, fields, strings, space-time points, or something similar.

each fan, so the qualitative character of my mental state is constituted by the qualitative character of the items that make up the relevant part of my brain. While the ontology of the physical sciences does not include phenomenal properties, the *explanatory shape* of the physical sciences—viz., a broadly bottom-up one—carries over to phenomenal properties. There are no ontic gaps in the hierarchy of nature; phenomenal properties do not constitute "nomological danglers."[14] The *most* elegant version of panpsychism is what Chalmers calls "Russellian constitutive panpsychism." It is the view that phenomenal properties are the categorical bases of physical properties—e.g. rest mass, negative charge, gravitational force, etc.—properties which physics describes wholly in terms of their dispositional profiles.[15] This view can explain not only where phenomenal properties come from, but why there are as many basic types of phenomenal property as there are. Chalmers contends that a further virtue of the Russellian version of panpsychism is that it gives phenomenality a causal role to play in the universe. (Whether it gives phenomenality the *right* causal role isn't so clear.)

The central challenge for panpsychism is to explain the constitution-relation that is said to hold, with metaphysical necessity, between the phenomenal properties of the UPCs and the phenomenal states of macro-level minds. This is the so-called "Combination Problem" for panpsychism. There are many aspects to the problem, corresponding to various explanatory questions that can be raised in connection with the

---

[14] The phrase comes from Smart (1959).

[15] Alternatively: phenomenal properties *just are* these properties: physics individuates them indirectly, in terms of their causal roles, and it turns out that phenomenal properties are the realizers of those roles. This is a kind of physicalism about phenomenal properties, though not in the sense of 'physicalism' I favor.

constitution-relation in question. I will briefly list four such sub-problems, and then

discuss what I take to be the best strategies for solving them.[16]

(1) *The Unity Problem*: How do the many phenomenal properties of the UPCs

form a single, unified the conscious state?

(2) *The Palette Problem*: How do the phenomenal properties of the UPCs

generate the range of types of phenomenal property, and the abundance of

determinates thereof?

(3) *The Selection Problem*: How are the phenomenal properties of the UPCs

differentially "selected" so as to contribute to a dynamic, evolving conscious

state?

(4) *The Structure Problem*: How do the phenomenal properties of the UPCs form

a conscious state with multi-modal structure, semantic structure, visual-field

structure, attentional structure, and so forth?

A reason to think that the Unity Problem cannot be solved is that panpsychism can be

made the target of the same sorts of Cartesian arguments that physicalism can: it seems

that no matter how many little consciousnesses are in the picture, it is a *further fact* that

those consciousnesses combine to form a big consciousness (if they do). Philip Goff, who

used to be convinced by such reasoning,[17] has recently changed course and concluded

that there must be a hitherto unknown element that functions to unify distinct

phenomenal states—an element such that, if we understood its nature and knew where it

was to be found in nature, would render perspicuous the connection between little minds

---

[16] My discussion overlaps with and is informed by Chalmers (ms).
[17] Goff (2009)

and the big minds they compose.[18] He calls this element "phenomenal bonding."

Chalmers (ms) builds on the proposal, suggesting (a) that the phenomenal bonding relation might be identical to the co-consciousness relation (the relation that holds, e.g., between phenomenal states of distinct modes, thereby rendering them states of the same mind), and (b) that it might be assimilated within Russellian panpsychism by identifying it with the intrinsic nature of one or another *relation* posited by physics—e.g. causation or spatiotemporal location.[19] One obvious problem with this proposed identification is that it would render phenomenal bonding—and hence the occurrence of macro-level consciousness—utterly ubiquitous. Perhaps this consequence could be embraced, though I think it would count as a liability of panpsychism if it *must* be embraced. Notice that not only would we be forced to grant that books and omelets and oceans have interior lives, but we would be forced to grant that our own subjective experience overlaps with the experience of innumerable subjects distinct from ourselves. We would be faced with a bewildering "problem of the many" for conscious states.[20] Because emergentism faces parallel problems (which I take up below), I won't pursue the matter further here. But the problem is a serious one and I am not sure panpsychists will be able to solve it.

The Palette Problem is only a problem if the range of properties that make up our conscious experience outstrips the range of properties instantiated by the UPCs. On Russellian panpsychism, this certainly seems to be the case: if phenomenal properties are taken to be the categorical bases of properties posited by fundamental physics, then there

---

[18] Goff (forthcoming).

[19] Goff (forthcoming) toys with the stronger claim that the phenomenal bonding relation *just is* the spatial-distance relation—though not for Chalmers' Russellian reasons.

[20] The "problem of the many" is so named by Unger (1980). Note also that we would lose one of the motivations discussed in chapter 2 for the Phenomenal Grounding Thesis, viz., the need for a criterion for the individuation of cognitive systems. If every system is conscious, then consciousness cannot give us a useful way of individuating cognitive systems.

can only be as many phenomenal properties as there are fundamental physical properties. If we treat every point in each phenomenal quality-space as a distinct phenomenal property, then we will be left with *countless* phenomenal properties. One obvious move here is to suggest that there need only be as many *determinable* phenomenal properties as there are fundamental physical properties. The idea might be that distinct magnitudes of a fundamental property correspond to distinct determinates of a phenomenal property. There are problems with this idea, however: colors seem to vary along three dimensions (hue, saturation and brightness) whereas we are not aware of any physical properties that admit of such structures; further, as Chalmers (ms) points out, not all fundamental physical properties admit of scalar magnitudes of any sort. An alternative move would be to give up on Russellian panpsychism and instead to allow that the number of types of phenomenal property instantiated by UPCs is very large. Of course, the number of UPCs in the brain of a human subject is also very large. So long as at least one of the UPCs in my brain instantiates *this* determinate shade of phenomenal green (according to this way of thinking), then that determinate shade can show up in my conscious experiences. This strategy for solving the Palette Problem comes at significant loss of theoretical elegance. Phenomenal properties would turn out to be arbitrarily distributed across the UPCs; there would be no explaining why any given UPC instantiates *this* phenomenal property and not another.

A superior strategy than either of those just discussed is to say that the phenomenal properties instantiated by the UPCs form a sparse set, and that these properties jointly generate all the phenomenal properties of our acquaintance. An example of this approach is the "phenomenal blending" strategy of Luke Roelofs (2014).

Roelofs proposes that phenomenal properties can "blend" to form new phenomenal properties. Illustrations come from cases in which an aesthetic novice treats certain phenomenal properties as simple but that a connoisseur can recognize as complex: the sound of a musical triad or the taste of wine, for example. Roelofs suggests that constituting-properties of the UPCs might not be identical with any of the constituted-properties of our acquaintance. They might instead be "alien" properties we are not in a position to imagine. Now, we do instantiate these alien properties; that they blend does not mean that they dissolve into the blend (which would probably amount to a kind of emergence). Nevertheless, we cannot *imagine* them because we lack the capacity to instantiate them *in isolation*, unblended.

So far, so good: by bringing phenomenal bonding and phenomenal blending into the picture, the panpsychist has at least the sketch of a solution to the Unity Problem and the Palette Problem. But the Selection Problem and the Structure Problem remain. Consider the obvious fact that one's consciousness is populated by different phenomenal properties at different times. The Selection Problem is the problem of understanding what could explain this: why a subset of the available UPCs contributes to a unified, phenomenally bonded state at a time, and a different subset at a different time. The Unity Problem is solvable insofar as phenomenal bonding is pervasive; the Selection Problem is solvable insofar as phenomenal bonding is *restricted*. Similarly: consider the obvious fact that one's consciousness contains structure of various sorts (multi-modal, semantic, visual-field, attentional, etc.). The Structure Problem is the problem of understanding what could explain this: why the phenomenal properties that result when the UPCs are suitably related to one another clump together in various ways and not in others. The

244

Palette Problem is solvable insofar as phenomenal blending is pervasive; the Structure Problem is solvable insofar as phenomenal blending is restricted.

What would an explanation of such restrictedness (of bonding and blending) look like? One sort of explanation would be in terms of the absence of a catalyst for bonding or blending; another would be in terms of the presence of an inhibitor of bonding or blending. Either way, we would still need an account of why and when such a catalyst or inhibitor gets deployed, and such an account will need to reference the *neural-functional conditions* that correspond to certain experiences. We know, for example, that particular visual experiences are occasioned by retinal stimulation followed by activation of particular regions in visual cortex; particular auditory experiences are occasioned by stimulation of inner-ear structures followed by activation of particular regions in auditory cortex, and so forth. We know, in other words, that which type of experience a subject is having is a matter of which processing-streams in the brain are active. We know, too, that these sorts of processing-streams are multiply realized. First of all, there is considerable variation in how a processing-stream is implemented, both across brains and in the same brain at different times (in stroke patients, for example). Second, functionally identical neurons can be made of different types of organic molecule.[21] Neuroscience is thus confirming what early functionalists such as Putnam (1967) and Fodor (1974) were urging several decades ago: the instantiation of certain *functional* states is necessary for the instantiation of certain mental states. These functional states are not identical with physical states, nor even with brain states, since they are realized in multiple ways at both the physical and the neural level. This is going to be a problem for any "bottom-up"

---

[21] Aizawa & Gilet (2009).

approach to mental states—for panpsychism just as much as for type-identity theory. (Also for certain forms of emergentism—as we'll see in section 4 below.)

What this means for panpsychism is that whether a UPC participates in bonding/blending at a time is a matter of what is going on neural-functionally. Any specification of the limits of bonding/blending must invoke either (a) high-level, multiply realizable neural functions; or (b) all the possible physical conditions under which those neural functions are realized. Either way, there is no smooth, bottom-up explanatory relationship between the phenomenal properties of the UPCs and the unified phenomenal state those properties combine to form. And if option (b) is off the table (as I think it should be; "explanations" in terms of massive disjunctions are not really explanations), there is *no* bottom-up explanatory relationship between the phenomenal properties of the UPCs and the unified phenomenal state those properties combine to form.

Here is one way that a panpsychist might respond: yes, neural-functional properties are multiply realizable, high-level properties; and yes, which properties of the UPCs bond/blend at a time is a matter of which neural-functional states obtain at a time. But neural-functional states are not the most proximal conditions for bonding/blending. The most proximal conditions for bonding/blending have to do with the states of a centralized "hub" of brain activity. This brain-hub could consist, at a minimum, of one representative UPC for each basic phenomenal property. (If Russelian panpsychism is true, then this will amount to exactly as many UPCs as there are fundamental physical properties.) Each of these UPCs stands at the end of a neural pathway. When a signal is transmitted along a neural pathway, its culminating UPC is stimulated, whereupon the phenomenal property of that UPC enters into bonding/blending. Hence the *distal*

conditions on bonding/blending involve multiply realizable, high-level properties, but the *proximal* conditions involve only properties at the physical level.[22]

The problem with the "brain-hub" hypothesis is that it is very likely false. Distinct neural pathways—such as the "color" pathway and the "motion" pathway in the visual system—do not converge on single spot in the brain, but rather remain almost entirely functionally isolated from one another. (Indeed they do not even process information synchronously.)[23] There is little consensus regarding the neural correlates of consciousness,[24] but most theories invoke large functional swaths of the brain, such as "thalamo-cortical loops."[25] Of course it's conceivable that we could find out that there is, after all, such a brain-hub, changes in which directly determine the features of phenomenal states.[26] But our best empirical findings do not indicate as much, so I will proceed on the assumption that the brain-hub hypothesis is false.

We should grant, then, that multiply-realizable, high-level properties *are* directly relevant to which phenomenal properties are included in a macro-level phenomenal state. But irreducibly macro-level influence is a traditional hallmark of emergentism! In short, any adequate panpsychism—one that can solve the Selection and Structure Problems—*is a form of emergentism*, albeit a uniquely panpsychist form of emergentism. Now, this is not itself to provide an argument *against* panpsychism. But it is to undercut much of the dialectical advantage of panpsychism over emergentism.

---

[22] Descartes' famous "pineal gland" hypothesis is very similar to the suggestion I'm exploring here.
[23] Zeki (2003).
[24] Chalmers (2000).
[25] Llinás (2003).
[26] We could then speak of "brain-hub-in-a-vat" thought experiments: a brain-hub is kept alive and stimulated in a manner identical to some normally enskulled brain-hub, thus resulting in two qualitatively identical streams of consciousness.

Matters are worse for panpsychism once P-I properties are in the picture. P-I properties pose a unique challenge to panpsychism in three ways. First, P-I properties render the Palette Problem harder to solve. This is because P-I properties do not bear the same relationship to one another that S-S properties do. S-S properties in a single mode can plausibly be located within a quality-space. We can render intelligible the idea that phenomenal-color properties (say) are the product of blending, to the extent that we have a grip on the dimensions along which phenomenal-color properties can vary. (Analogies with paint-mixing help, too.) But it is not clear how to locate P-I properties in a quality space. How are P-I properties whose intentional objects are (a) causation, (b) egocentric location, and (c) triangularity to be located with respect to one another in a P-I quality-space? It is no clearer how to do this than how to locate causation, egocentric location, and triangularity themselves in a quality-space. In consequence we have not the faintest idea what it could mean for P-I properties to be the result of blending. So the panpsychist probably should include all primitive P-I properties among the basic ingredients for blending. Depending on how many primitive P-I properties there are, this may require giving up on Russellian panpsychism. And whether or not it requires that, it certainly requires a loss of elegance. Instead of a few, alien properties had by the UPCs which together compose the phenomenal properties of our acquaintance, we must add to these *a seemingly arbitrary set of P-I properties of our acquaintance*. This is not a theoretically happy addendum.

Second, P-I properties render the Structure Problem harder to solve. This is because phenomenal bonding—the production of a unified phenomenal state out of discrete phenomenal states—and phenomenal blending—the production of various

248

phenomenal qualities out of a set of primitive, perhaps alien qualities—does not amount to phenomenal *binding*, the sort of metaphysical structure that accounts for *semantic* structure. I said above that the Structure Problem amounts to the problem of explaining why binding is *restricted*. But the Structure Problem is more than that, since the binding of two phenomenal properties is more than their non-blendedness. If my account of binding is correct, according to which P-I properties such as P-I attribution point beyond themselves toward other phenomenal properties, then the panpsychist will needs to supply an explanation of where such properties come from (presumably the likes of P-I attribution will also need to be instantiated by the UPCs) and why they structure phenomenal states in certain ways rather than others. Perhaps an ad hoc story could be told here. Again, this would not be a theoretically happy addendum.

The third challenge has to do with the possibility, discussed at the end of the last chapter, of new intentional primitives' emerging at various stages of cognitive development. If P-I properties are not the product of blending, then all primitive P-I properties need to be included among the "ingredients," so there cannot be "new" primitives (if new primitive P-I properties are *non-basic* P-I properties.) The panpsychist could say, perhaps, that P-I properties can be *developmentally* new without being *ontologically* new. Perhaps, that is, certain P-I properties *are* in the base, but are prevented from being included in macro-level phenomenal states until the cognitive system reaches a certain level of complexity. The viability of this suggestion will turn on how common new primitives are. If there are only a handful (as Susan Carey's work might suggest), then perhaps the suggestion is viable. But the suggestion looks unworkable if other types of "new" primitives I discussed are also included in the base:

P-I properties whose intentional objects are high-level sortals such as parts of speech, the contents apprehended during episodes of linguistic experience and/or purely cognitive experience, and derivative determinables such as sensation-types and spatial-types. The panpsychist would thereby be committed to a truly enormous base.

Taking stock: the appeal of panpsychism consists in its ability to explain how a small palette of phenomenal qualities combine, in familiar, bottom-up fashion, to form the macro-level phenomenal states of our acquaintance. Solving the Selection and Structure Problems means giving up on bottom-up explanation: high-level facts about neural pathways partly determine when bonding/blending occurs. Accounting for P-I properties means (a) expanding the base of micro-level phenomenal properties astronomically (to account for all the possible primitives), as well as (b) punting on, or at best complicating in ad hoc way, the explanatory relation that holds between the base-properties and the states they compose (to account for semantic structure).

In short, panpsychism bears all the marks of an initially promising but ultimately misguided theory. It is initially promising because it purports to draw explanatory parallels between the metaphysics of phenomenal states and the metaphysics of material composition. We have a good idea of how the physical and functional structure of material composites such as ourselves is determined by the dynamics of the UPCs that make us up; we know that material composites such as ourselves have a conscious "interior"; it stands to reason that this conscious interior is likewise determined by the properties of the UPCs that make us up. But closer examination does not bear this out: the dynamics of phenomenal states correlate with the dynamics of *mid-level* rather than *ultimate-level* parts of ourselves, and it is quite mysterious how the ubiquitous *intentional*

properties of consciousness could be made out of anything non-intentional. The appearances cannot be elegantly saved. If we have an alternative theory, we should explore it.

Some have suggested that emergentism is no alternative theory but rather a restatement of the phenomenon for which we need a theory: *once there are brains, consciousness appears on the scene*.[27] So saying is obviously correct but uninformative (or so goes the criticism). But I think that emergentism can be rendered informative once the connection between the emergence-base and that which emerges from the base is understood in *causal terms*.[28] Causation, I will assume, is a fundamental aspect of the world. It occurs when causal powers are manifested. A causal power is a dispositional property whose nature consists entirely in (a) its proprietary manifestation, i.e. the effect at which it aims, in connection with (b) the conditions under which it manifests. *Fundamental* causal powers are those causal powers that are not reducible to, realized in, or composed by any other causal powers. These causal powers are the ultimate explainers of the causal dynamics of the world. In rare cases, fundamental causal powers manifest in an isolated fashion, e.g. in cases of radioactive decay. Typically, many instances of fundamental causal powers manifest jointly; this is what is going on in collision-mechanics of any degree of complexity. Joint manifestation can result in cancelation, where the resulting effect is not that at which any of the powers is aimed. Any static physical system is the result of such cancellation. But joint manifestation can also resulting in amplification, where the resulting effect is greater than that at which any of the powers is aimed. Any macro-level motion is the result of such amplification.

---

[27] Cf. Nagel (1979), Bennett (ms).
[28] Here I follow O'Connor & Wong (2005).

Emergence is the result of a special sort of joint power-manifestation. Consider, by way of analogy, the difference between my ability to contribute causally to the lifting of a car, on the one hand, and to the formation of a club, on the other. The first is an example of the ampliative, joint manifestation of (non-fundamental) causal powers. That is, if I coordinate my efforts with others, we together have the power to lift a car, because each of us has the power to lift a part of the weight of the car. My contribution to the formation of a club is not quite like this, though. If I coordinate my efforts with others and we together form a club, this is not because I have on my own the power to form a part of a club. Instead, whatever power I have to form a club is *essentially* collective. It is not just a jointly manifested power, it is a jointly manifested *collective* power. Returning to the matter at hand: the emergence of consciousness is a matter of the joint manifestation of a collective consciousness-generating power had by the UPCs.[29]

Natural questions to ask at this point are: *how* and *when* does it happen? These questions raise a challenge for emergentism akin to the Unity Problem for panpsychism: just as panpsychism requires that many phenomenal properties combine to form unified phenomenal states, so emergentism requires that many UPCs causally contribute to the generation of unified phenomenal states. Now, these two requirements are a bit different: on emergentism, UPCs need not stand in a relation that unifies their properties with one another, but rather they must stand in a relation that allows them to coordinate their causal powers. It would perhaps be better to call the relevant explanatory problem for emergentism "the Coordination Problem" rather than "the Unity Problem."[30] The two

---

[29] The view I sketch in section 4 revises this picture slightly.

[30] Questions may remain for both panpsychism and emergentism regarding the nature of phenomenal unity. What panpsychism requires in addition, but that emergentism does not, is an account of *unification*, i.e. the mechanism whereby non-unified phenomenanl properties come to be unified.

252

questions just mentioned—the *how* question and the *when* question—amount to two aspects of the Coordination Problem. The first question is the question of what makes it possible for UPCs to coordinate their causal efforts. The second question is the question of why such coordination occurs only in special circumstances. I take these two aspects of the Coordination Problem to amount to the most difficult challenges for emergentism. I address them in turn.

How does coordination occur? We now need to confront some disanalogies with the club-formation case. When I join with others in forming a club, what makes possible the joint manifestation of our collective power is a certain shared intention to do so, and this shared intention is possible because of communication among us. I do not want to say that an aggregate of UPCs can share an intention, nor even that they can communicate with one another, in anything like the sense of 'communicate' relevant to the club-formation case. So some other account is needed of the nature of coordination among the UPCs. And I think that this account has to supply a pretty strong form of unity among the UPCs. Precisely because the causal power in question is a *collective* power, the coordinated UPCs have to act *as one*; they are not each generating a part of phenomenal states but rather are jointly, synchronously generating entire phenomenal states. Fortunately, the idea of spatially separated entities acting as one is not an idea totally foreign or repugnant to contemporary physics (as it would have been 150 years ago). Quantum-entangled entities can act as one, even when separated by great distances. I tentatively speculate that quantum-entanglement in the brain is what makes

coordination possible.[31] (If this speculation proves unworkable, some heretofore

unknown unification-relation will need to posited in its place.)

This cannot be the whole story, assuming that some entangled systems are not

conscious. So there must be an additional condition on the manifestation of

consciousness-generating powers. What is that condition? This is a hard question to

answer, without some sense for the *scope* of consciousness. Philosophers and cognitive

scientists are far from finding a consensus here.[32] One rough and roundabout way to look

for an answer is to ask after the *function* of consciousness, i.e. the way that consciousness

contributes toward the fitness of systems that give rise to it. An intuitive proposal is that

consciousness economizes the mechanisms driving flexible behavior.[33] Here is the idea:

it's easy to get a robot to exit the door. It's much harder to get a robot to go out the door

and go to the grocery store. Why? Because at each step of the way, alternative

possibilities for behavior must be ruled out mechanically. Much simpler would be for the

robot to *want to go to the store*. How much simpler, that is, if the robot could entertain

explicit goals and feel motivation with respect to those goals. I suggest, therefore, that the

most basic function of consciousness is to motivate goal-directed behavior. But of course

this function is of no help to a system that has no sense apparatus or motor skills.

Consciousness is of service, in the first instance, to a system that implements the

equivalent of a sensorimotor system.[34] So I propose the following answer to the 'when'

---

[31] The idea that quantum phenomena in the brain are relevant to consciousness is somewhat new, but is becoming more widely explored. See Hameroff & Penrose (2014) and Craddock, et al (2015).
[32] See Allen & Triesman (2015).
[33] Compare Merker (2005) and Trestman (2013a) & (2013b). I discussed the notion of flexibility in more detail in chapter 2.
[34] McGinn (1988) articulates a similar idea: "First there were preconscious states with certain functions relating them to things in the world; then consciousness built upon this natural foundation to produce the intentional relation. The 'intentional arc' is not reducible to this foundation but it takes its rise from it" (p. 242).

question: UPCs manifest their consciousness-generating powers when they form a sensorimotor system, or a sensorimotor system above a certain threshold of complexity.

Now, this criterion might seem hopelessly vague (or if not vague, then arbitrary). What counts as a sensorimotor system? Does a robot programmed to randomly move about without running into things count? I am not sure what to say here, but I don't think the emergentist is out of options. For example, it is known that complex systems can undergo dramatic, sudden transitions when they reach certain levels of complexity.[35] So such "non-linear transformations" very well could supply non-arbitrary criteria for the emerging of consciousness.[36]

My answer to the *when* question—in terms of sensorimotor systems—amounts to a fairly liberal criterion for the emergence of consciousness. Insects and simple robots can implement sensorimotor systems, after all. Note, though, that my answer to the *how* question—in terms of entanglement—supplies a separate criterion, and the two criteria are, presumably, dissociable: some sensorimotor systems might not have entangled parts, and some entangled systems might not be sensorimotor (I take it this latter dissociation happens all the time). I have said nothing about the *explanation* of entanglement. Perhaps brain-embedded UPCs become entangled precisely because they compose a sensorimotor system, in which case the two criteria are not fully dissociable after all. But perhaps brain-embedded UPCs become entangled as a result of some other feature of brains. And

---

[35] Kelso (1995).

[36] One of the motivations discussed in chapter 2 for the Phenomenal Grounding Thesis pertained to the problem of individuating cognitive systems. It might be argued, on the basis of my discussion here, that whatever criterion the emergentist uses to sort conscious systems from non-conscious systems can be used by the functionalist to sort cognitive systems from non-cognitive systems. Perhaps, but the functionalist would not thereby have escaped arbitrariness problems. The emergentist says that at certain levels of complexity, systems are caused to undergo an ontological change. The functionalist says nothing of the sort. So the functionalist remains unable to explain why her criterion is metaphysically significant.

if this feature is *unique* to brains, or to large brains, or to carbon-based systems, then my criterion may turn out to be less liberal than it initially appears.

It looks as though we can at least get emergentism up and running, then. Given the problems that have mounted for panpsychism, emergentism appears worth taking seriously. A well-founded adjudication between the two views can only be reached once I have subjected emergentism to the same level of scrutiny as I have subjected panpsychism. I'll be doing exactly that in the remainder of this chapter, as I continue to develop a version of emergentism consistent with my theory of P-I properties.

3. Second Choice Point: Emergent Materialism vs. Emergent Dualism

A central question that comes up for emergentism (but that does not come up for panpsychism) is this: what is the *bearer* of phenomenal states? The more ontologically conservative answer is *emergent materialism*, according to which the bearers of emergent phenomenal properties are composites of UPCs. The more ontologically profligate answer is *emergent dualism*, according to which the bearers of emergent phenomenal properties are ontically novel simples (i.e. immaterial Souls[37], or Selves[38], or Subjects[39]). Given that emergent dualism is a form of substance dualism and hence considerably less parsimonious, we should opt for it only if emergent materialism proves theoretically inadequate. I will discuss two recent attempts to show that emergent materialism is in fact inadequate, from Dean Zimmerman and William Hasker. I reach two conclusions. First, a theoretically adequate emergent Materialism can escape their criticisms, so their arguments do not succeed. However, a closely related argument succeeds in pushing

---

[37] Swinburne (1997), Zimmerman (2010).
[38] Hasker (1999).
[39] Nida-Rumelin (2007), LaRock (2013).

emergent materialism in an inflationary direction. It turns out (this is the second conclusion) that emergent materialism does not enjoy the major advantage over emergent dualism, with respect to parsimony, that it is often taken to enjoy.

Zimmerman (2010) argues that all material candidates for being the bearers of emergent phenomenal properties are problematically vague. Given emergentism, the instantiation of phenomenal properties—or 'qualia,' as Zimmerman prefers to call them—is not necessitated by the laws of physics, but must be governed by "fundamental laws of qualia generation." These laws specify (a) the conditions under which a quale is generated, (b) which particular quale is generated, and (c) which object instantiates it. Hence the bearers of qualia must be mentioned in those laws that amount to the explanatory grounds of their emergence. But which objects are these? Here a dilemma opens up for the emergent materialist. On the one hand, they could be what Zimmerman calls "Garden-Variety Objects," i.e., organisms or parts of organisms such as brains or central nervous systems. But all of these objects have vague spatiotemporal boundaries; there is no fact of the matter about where/when they begin and end. And this is a problem, says Zimmerman, because *fundamental laws don't mention vague objects*. On the other hand, qualia-bearers might be none of these familiar, vague objects, but instead be unfamiliar, sharp objects. But how are we to decide between the many, sharp objects that partly overlap the vague ones? It is hard to imagine a metaphysically respectable criterion. Let's call this "The Argument from Vagueness." Put more formally:

(AV1) Candidates for being the material bearer O of an emergent phenomenal state Q of are either familiar, vague objects or unfamiliar, sharp objects.

(AV2) If O is a familiar, vague object, then fundamental laws make mention of vague objects.

(AV3) If O is an unfamiliar, sharp object, then there is a non-arbitrary criterion for delineating their boundaries.

(AV4) Fundamental laws do not make mention of vague objects.[40]

(AV5) There is no non-arbitrary criterion for delineating the boundaries of the relevant sharp objects.

(AV6) Hence, qualia-bearers are not material.

A second argument comes from William Hasker (ms). Hasker is interested in the apparent mismatch between the unity of a phenomenal state and the multiplicity of UPCs that make up a material composite. How is a unified phenomenal state instantiated by a multiplicity of constituents? One option is that the state, Q, is instantiated by the material composite, O, in virtue of *parts* of Q being instantiated by *constituents of* O. But Hasker takes it as given that this is impossible. Just as your instantiating a phenomenal state and my instantiating a phenomenal state could never combine to form a third, joint phenomenal state—your state remains forever privately yours and mine forever privately mine—so Q could never be the result of some sort of combination of the phenomenal properties instantiated by the various UPCs one by one. Thus Hasker seems to take the unitary nature of phenomenal states to imply that such states are not mereologically

---

[40] In conversation, Zimmerman has told me that he is not committed to laws' being explanatorily bedrock. Laws could be explanatorily derivative off of causal powers. Perhaps the argument could be recast in terms of the specification of the manifestation of the relevant powers—in particular, that such a specification will mention not only mention the generation of qualia but also the *bearer* of the generated qualia. Premise X could then be replaced with the claim that the specification of the manifestation of fundamental powers does not invoke vague entities.

composed.[41] (And he is right to do so: explaining the mereological composition of phenomenal states would amount to a Combination Problem for emergentism. Emergentism would thus inherit the liabilities of panpsychism while leaving the virtues of panpsychism behind.) A second option is that Q is instantiated by O in virtue of Q's being instantiated wholly by every constituent of O. But this would mean that when Hasker is enjoying Beethoven's Ninth Symphony, so is each quark in his brain (or whatever object O is supposed to be)—and Hasker finds that this idea "strain[s] one's credulity to the breaking point, and beyond."[42] A final option is that Q is instantiated by O as a whole but not in virtue of the properties of its constituents: it is "spread out," as it were, all over O, and only over O. But if this were so, Hasker reasons, then not only would the properties that make up Q fail to be found in every proper part of O, they would fail to be found in *any* proper part of O. And among the proper parts of O is the fusion of all of its constituent UPCs *save for one quark*. The implication would be that in O as a whole, but not in O minus one quark, the properties that make up Q are to be found. Hasker does not provide an argument for why he rejects this result, but perhaps his reasoning is similar to Zimmerman's: making such a cut-off will be metaphysically arbitrary; nature simply doesn't supply a criterion for doing so. The consequence is that

---

[41] Matters are complicated here, because phenomenal states clearly *have* parts: they can consist in part in perception and part in conscious thought; they can consist in part in pleasant sensations and part in painful sensations; perhaps they can be partly representational and partly purely qualitative. I gather that Hasker would grant all of this but be unmoved in his core intuition that phenomenal states cannot have actually existing, numerically distinct phenomenal states as proper parts (or at least that such parts could be metaphysically prior to the whole). Panpsychists will deny Hasker's core intuition. The reason it is not dialectically otiose for Hasker to rely on his core intuition in the present context is that his disagreement is with fellow emergentists. The most common reason (though by no means the only one) to find emergentism preferable to panpsychism is because panpsychism requires just this sort of mereological combination of phenomenal states.

[42] Hasker (ms), p. 39.

there is no way for a composite such as a brain to instantiate Q. Call this "The Argument

from Compositeness." Put more formally:

(AC1) If a composite object O instantiates an emergent phenomenal state Q, it does

so either because O's constituents instantiate (a) all of Q, (b) parts of Q, or (c)

none of Q.

(AC2) UPCs do not instantiate phenomenal states so rich as ours, so (a) is false.

(AC3) Phenomenal states are not mereologically composed, so (b) is false.

(AC4) There is no non-arbitrary criterion for delineating the boundaries of O (i.e.

(AV5) above), so (c) is false.

(AC5) Hence the bearers of emergent phenomenal states are not composite.

Hasker reaches the further conclusion that the bearers of phenomenal states are not

material. And this does follow, given that (per (AC3)) material non-composites such as

UPCs cannot be the bearers of full-blown phenomenal states.

Both of these arguments can be resisted, in light of the account of emergence

developed in the previous section. Contra the premise shared by the two arguments,

(AV5) and (AC4), the material object O that instantiates emergent phenomenal states has

a sharp boundary, and there is a non-arbitrary criterion for delineating this boundary.

Here is why. Central nervous systems may very well be vague objects. There are UPCs

such that it is vague whether they are part of my CNS. It will not, however, be vague

which of the UPCs are contributing at any one time to the joint production of my

consciousness. This is true even if more of the UPCs contribute to the manifestation than

are required to do so. (Compare: the club could have been formed if one of its founding

members had not shown up for its founding. Nevertheless, all of those who contributed to

260

its formation are part of it, at least at the time of its formation.) I suggest that the composite whose parts are all and only those UPCs that contribute to the joint generation of a phenomenal state at a time is the bearer of the phenomenal state at that time. *The psychological subject* is to be identified, at a time, with that material composite coincident with the set of UPCs causally responsible for the generation of that subjects' consciousness at that time. Emergent Materialism thus survives the criticisms of Zimmerman and Hasker. We can embrace the second horn of Zimmerman's dilemma: O, the bearer of emergent phenomenal state Q, is sharply-bounded. And we can embrace the third horn of Hasker's trilemma: that O instantiates emergent phenomenal properties does not imply that any proper parts of O instantiate emergent phenomenal properties.

Still, there are reasons in the neighborhood of those adduced by Zimmerman and Hasker to worry that my account is unsatisfactory. The concern I have in mind is whether O, sharply-bounded though it may be, is the right sort of thing to be the bearer of fundamental properties. O is what we might call a "loose composite": it is an aggregate of UPCs that more or less cohere with one another. Call a "basic bearer" of a property something whose instantiating of a property is not in virtue of any other property-instantiation. (If O instantiates Q, but it does so not because any of O's parts instantiate Q or part of Q, then O is a basic bearer of Q.) There is reason to wonder whether loose composites can be basic bearers of fundamental properties.

Suppose there are three UPCs, A, B, and C, scattered throughout the universe, but exhibiting the following peculiar commonality: wherever one is found, there also is found an instance of a fundamental property—a color, let's say—or better, an instance of the dynamic unfolding of a *sequence* of colors. Let's call the fusion of A, B, and C 'Comp'.

Suppose we were asked to specify the basic bearer or bearers of the color-sequence. Should we say that there are three basic bearers (A, B, and C), or that there is one basic bearer (Comp)? Without any more information about the case, there is no reason to say that there is one basic bearer rather than three—and if there is no reason to say so, then there is nothing to *make* it so. If Comp instantiates the color-sequence, it does so derivatively. Now, couldn't we say in response: *there's just a fact of the matter* as to whether there is one basic bearer or three? I have no conclusive argument against saying this. But neither is it satisfying. Here as elsewhere, we should privilege bottom-up explanatory structure. That is, we should assume that properties of composites are determined by properties of their parts. If bottom-up explanatory structure breaks down— as it would, were Comp a basic bearer of the color-sequence—we ought to require an *explainer* of this breakdown.

If this is right, it spells trouble for the version of Emergent Materialism I have so far sketched. The UPCs that constitute my CNS are also "scattered," if only just barely: the fact that they are in close spatial and causal proximity doesn't change their status as an aggregate, as composite in only a *loose* way. Does the fact (if it is a fact) that they are *entangled* change their status as an aggregate? I don't see that it does. Entangled UPCs may *act as one* but this is not to say that they have literally *become one*. So we should reach the parallel conclusion regarding the bearer of emergent phenomenal properties: when UPCs conspire to jointly generate phenomenal properties, there is no good reason to say that the generated properties are instanced only once, with the relevant *composite* as their bearer, rather than that they are instanced many times over, as many times as

there are UPCs that form the composite. (With Hasker I treat this as a *reductio ad absurdum* of Emergent Materialism.)

So even if there is nothing objectionable about the very idea that a composite could instantiate a phenomenal property without any of its parts instantiating that property (or any part of that property), there is a closely related idea that *is* objectionable, viz., that a mere aggregate could instantiate a fundamental property, without any of its parts instantiating that property (or any part of that property).[43] If we swap this point in for premise 4 in Hasker's argument, we can construct a new argument, which I'll call "The Revised Argument from Compositeness."

(RAC1)   If a **loose** composite object O instantiates an emergent phenomenal state Q, it does so either because O's constituents instantiate (a) all of Q, (b) parts of Q, or (c) none of Q.

(RAC2)   UPCs do not instantiate phenomenal states so rich as ours, so (a) is false.

(RAC3)   Phenomenal states are not mereologically composed, so (b) is false.

(RAC4)   Loose composites are not basic bearers of fundamental properties (such as those that make up Q), so (c) is false.

(RAC5)   Hence the bearers of emergent phenomenal states are not **loose** composites.

The Revised Argument from Compositeness is weaker than Hasker's original argument. The original argument sought to rule out the possibility that the bearers of emergent phenomenal states are composites. Insofar as material simples are not good candidates for being the bearers of emergent phenomenal states (per premise 3 in both arguments), it

---

[43] In conversation, Hasker has criticized what he calls the "magical holism" required by emergent materialists. So the criticism I am currently unpacking might be aptly attributed to Hasker, though it does not show up in his paper.

follows that the bearers of emergent phenomenal states are immaterial. But the revised argument seeks only to rule out the possibility that the bearers of emergent phenomenal states are *loose* composites. It remains open that the bearers of emergent phenomenal states are *strict* composites, i.e. true unities, bona fide individuals, despite having material parts.

What might a strict composite be? A strict composite must be more than a mere aggregate; it must include some element or elements that explain its deep, objective metaphysical unity. Note that providing an account of this element is not the same as providing the conditions necessary and sufficient for a strict composite to exist. (When, for example, Peter van Inwagen says that simples compose a new thing iff they jointly form a *life*, he is not *explaining* the unity of such composites; he is only telling us when and where to find such composites.) One account of strict composites is the 'Emergent Individuals' view of Timothy O'Connor and Jonathan Jacobs ([2003] & [2010]). O'Connor & Jacobs follow David Armstrong in holding that fundamental particulars have a complex structure, consisting of (a) one or more immanent universals, and (b) a "thin particular," i.e. an entity which particularizes universals when they inhere in that entity. So-structured fundamental particulars make up the world of our acquaintance, replete as it is with loose composites of all sorts. But under certain circumstances, composites themselves can come to have their own proprietary thin particular. Such composites are the bearers of emergent phenomenal properties, and are themselves the products of emergence: they are materially-composed *emergent individuals*.

I am not sure that the Emergent Individuals view of O'Connor & Jacobs is the only way to account for strict composites. Perhaps we need not invoke a special, new

thing (such as a thin particular) to do the job; perhaps certain new properties and relations could do the job instead. Perhaps, for example, there are special, contingent building-relations that hold between UPCs when those UPCs form strict composites. But I confess I am skeptical that such accounts could deliver the goods: shouldn't we just say in such cases that the UPCs form a comparatively tight-knit aggregate, rather than that, by anybody's reckoning including God's, a *new thing* has come to be?

There is no space here to follow these lines of inquiry. The broader point is this: emergent materialists can supply a viable metaphysics of the bearers of emergent phenomenal properties, but they must inflate their ontology a bit. And the result of the inflating is a picture *very similar in many respects* to that of the emergent dualists. Representatives from both camps can say the following:

> Psychological subjects depend on and emerge out of a physical aggregate, but are not identical to that aggregate. Psychological subjects are *fundamental* entities: they must be included even in the sparsest of inventories of what exists.

If the Emergent Individuals view is the right account of strict composites, we can add:

> The generation of psychological subjects involves the generation of an entity that is not materially composed [for the Dualist, a soul; for the Materialist, a thin particular].[44]

In short, emergent materialism may still beat emergent dualism when it comes to parsimony. But the margin of victory is much smaller than is usually supposed. Thus my

---

[44] Compare O'Connor & Jacobs (2010): "Particularity, as we conceive it, is in no sense a peculiar sort of property. It is, rather, a peculiar sort of particular: an entity that is incomplete in itself—its role is to particularize a complete object—and that in every case (plausibly) essentially instantiates certain properties and contingently instantiates others. (Granted, such entities can seem mighty peculiar indeed. But every account of objects and their properties has its peculiarities, and we judge this one to be a bit less peculiar, all things considered, than its rivals, and that is all that one can ask of a metaphysical theory.)"

defense of the *viability* of emergent materialism can also serve as an argument against the *wild implausibility* of substance dualism.

4. Third Choice Point: Bottom-Up Property Emergence vs. Top-Down Property Emergence

Here is the shape so far of our account of the emergence of phenomenal states. When UPCs that are parts of a sensorimotor system become entangled, they jointly manifest a collective consciousness-generating power. The immediate result is that these UPCs both (a) compose an emergent individual, and (b) generate a phenomenal state instantiated by that emergent individual.

This account does not answer all pressing explanatory questions for emergentism. Recall the Selection and Structure Problems for panpsychism:

The Selection Problem: How are the phenomenal properties of the UPCs differentially "selected" so as to contribute to a dynamic, evolving conscious state?

The Structure Problem: How do the phenomenal properties of the UPCs form a conscious state with multi-modal structure, semantic structure, visual-field structure, attentional structure, and so forth?

Very closely related questions come up for emergentism. They are not questions about how phenomenal properties *of the UPCs* are selected and structured so as to form the macro-level phenomenal states of our acquaintance. Rather, they are questions about how phenomenal properties *generated by the UPCs* are selected and structured so as to form the macro-level phenomenal states of our acquaintance. In other words, to say that UPCs

266

jointly manifest a collective consciousness-generating power is not yet to explain why

specific conscious states are generated. A more ontologically conservative answer is

bottom-up property emergence (or BUPE); a more ontologically profligate answer is *top-down property emergence* (or TDPE). According to BUPE, the specificity of emergent

phenomenal states can be explained in the same way as the emergence of consciousness

generally, viz., in terms of the powers of the UPCs. According to TDPE, the specificity of

emergent phenomenal states cannot be explained in terms of the powers of the UPCs; an

additional "top-down" causal element is required. I take it that BUPE is more

ontologically conservative than TDPE, for two reasons. First, bottom-up explanations are

more elegant than top-down explanations. (Thus BUPE has the same prima facie

advantage over TDPE that panpsychism has over emergentism.) Second, an explanation

in terms of one set of causes is more parsimonious than an explanation in terms of two

sets of causes. So we should go in for TDPE only if BUPE proves inadequate.

Here is the simplest version of BUPE. Each type of UPC is responsible for one

type of phenomenal property: one type of UPC contributes phenomenal blue; another

type contributes painfulness; and so forth. When a collection of UPCs generate a

phenomenal state, some of the UPCs in the collection contribute particular qualities to the

state. Hence UPCs and their powers are all that is needed to explain the emergence of

specific phenomenal states. (The most elegant form of this view would tie distinct

property-generating powers to distinct types of UPCs recognized in physics—the

different types of sub-atomic particle, say).

This view looks very much like panpsychism (and its most elegant form looks

very much like Russelian panpsychism); it differs only in that the relationship between

the properties of the UPCs and macro-level phenomenal states is *generation* rather than *constitution*. It is also subject to most of the same worries as panpsychism. Whether a UPC contributes its proprietary property on some occasion will depend in part on the neural-functional facts about the system in which it is embedded. So the manifestation-conditions for a UPC's property-contributing power will involve either (a) the obtaining of neural-functional states, or (b) one of a list of possible realizer conditions for such neural-functional states. We ought to avoid attributing essentially disjunctive manifestation-conditions to a fundamental causal power if at all possible. But then it looks as though this view cannot supply bottom-up explanations after all.[45] Further, in order to account for all of the possible determinates of phenomenal properties, the view will likely need to include some account of phenomenal blending. But since the idea of generating P-I properties via blending doesn't make sense, all P-I properties will need to be the proprietary causal products of a type of UPC—and this will mean either denying that new primitives are possible, or else including all such "new" primitives as basic causal products of some type of UPC. A third and final worry is that the view cannot explain phenomenal binding. Phenomenal states are more than a bundle of phenomenal properties: they admit of rich structure, including semantic structure. It is hard to see how the present proposal to explain this.

On an improved version of BUPE, individual UPCs do not contribute specific properties to phenomenal states. Rather, just as the generation of consciousness is an essentially collective effort, so the generation of the particular properties that make up states of consciousness is also an essentially collective effort. The idea here would be that

---

[45] The "brain-hub" hypothesis discussed in section 2 could serve to mitigate this worry, just as it could serve to mitigate the corresponding worry for panpsychism. But that hypothesis isn't empirically plausible.

specific, structured phenomenal states are brought into being in their entirety, as the result of the exercise of the consciousness-generating power *of the collection* of UPCs, in *its* entirety. And *which* specific, structured phenomenal state gets generated at a time is a matter of the causal relations among the UPCs. That is, the unique ways that the brain is activated at a time are relevant to which phenomenal state the brain generates at that time.

This version of BUPE helps to account for new primitive P-I contents: at certain stages of brain-development, the UPCs generate phenomenal states that include novel P-I properties. It also helps to account for phenomenal binding: since phenomenal properties are not generated piecemeal but rather are generated as part of whole phenomenal states, the structural elements of such states require no additional explanation. But problems related to multiply realizability remain. On the present proposal, collections of UMCs have the power to generate specific phenomenal states. We should probably understand this to mean that collections of UPCs have a single, "multi-track" power. (A multi-track power is a power that cannot be fully characterized by a single conditional of the form, "under conditions C, manifestation M occurs". The relevant power cannot be so characterized because it admits of lots of types of manifestation—as many types as there are possible phenomenal states for it to generate. Alternatively, we could understand collections of UPCs to have as many *powers* as there are possible phenomenal states for it to generate. But we should avoid multiplying powers in such fashion.) The problem related to multiple realizability comes in when we try to specify the conditions under which the power manifests in one of its many ways. The conditions under which brains generate a phenomenal states of a particular type are not happily described in the language of physics, but rather in the language of high-level, multiply realizable neural

functions. Again: we ought to avoid attributing essentially disjunctive manifestation-conditions to a fundamental causal power if at all possible. The general lesson is this: neural-functional states are relevant, as such, to the specificity of phenomenal states. Because neural-functional states are multiply realizable, no elegant, bottom-up explanation of the specificity of phenomenal states is forthcoming.

So we should go in for a version of TDPE. What would a "top-down" theory of the emergence of phenomenal properties look like? To start with, the bearers of phenomenal-property-generating powers would need to be at least as "high-level" as neural pathways, since it is goings-on in such entities that, as a matter of empirical fact, account for which properties are generated. A natural suggestion, then, would be to treat neural pathways as the bearers of the relevant powers. Nor is this an utterly counterintuitive suggestion: it is easy to think of the pathways in the visual system, for example, as taking sensory stimulation *as input* and as generating perceptual experiences (inter alia) *as output*—somewhat akin to the way a radio receives radio-waves as input and generates audible sound as output. But it is hard to work out the details of this suggestion. *Neural pathways* are odd entities—vaguely bounded, frequently morphing, constituted in part by the functions they implement. How are they individuated? Under what circumstances do they come to instantiate phenomenality-generating powers? The story is bound to be complicated; it may require treating neural pathways as emergent individuals in their own right (in addition to the emergent individual that serves as the bearer of the properties generated by the neural pathways). Furthermore, going this route would amount to a return to the "piecemeal" approach to the generation of phenomenal states. As we saw in connection with the first version of BUPE discussed above, a

270

piecemeal approach to the generation of phenomenal states doesn't seem to have the resources to account for phenomenal structure such as binding.

So I think we should bypass such "mid-level" approaches and go truly top-down: it is the emergent individual—hereafter the *emergent subject*—that is the bearer of a phenomenality-generating power. An emergent subject generates and instantiates phenomenal states, in response to states of the collection of UPCs that make it up. The states to which it is responsive can of course include "high-level" states such as neural-functional states. Thus problems related to multiple realizability do not come up for TDPE. Because emergent subjects generate whole phenomenal states, no problems related to structure come up, either. And the way that TDPE can accommodate new primitives is especially illuminating. At the end of the last chapter, I likened the grasping of new primitive contents to suddenly seeing a group of dots on a graph not as random but as delineating a line-of-best-fit. Something very much like this could occur at certain critical developmental moments: in response to certain activities in the brain, the emergent subject generates phenomenal states that involve rather complicated P-I contents; but when these P-I contents reach a certain critical threshold of complexity, a switch takes place, and a new primitive P-I content takes the place of what came before. This picture of the relationship between phenomenal states and brain states suggests a unique way of understanding the phenomenality-generating power of emergent subjects: it is very much like an *interpretive* power. Emergent subjects generate phenomenal states that amount to interpretations of the goings-on in the brain.[46]

---

[46] One might be puzzled by the suggestion that generating phenomenal states is something that subjects *do*. I am a subject; I do things such as think and act and perceive and so forth; *interpreting my brain* is not among these things that I do. Quite so, yet there *is* a sense in which interpreting my brain is something I do:

I previously said that collections of UPCs share a consciousness-generating power. When this power is manifested, the immediate result is that these UPCs both (a) compose an emergent individual, and (b) generate a phenomenal state instantiated by that emergent individual. We can now see that picture is not quite right. Collections of UPCs share a *subject-forming* power. When UPCs that are parts of a sensorimotor system become entangled, they jointly manifest their subject-forming power. The emergent subjects thereby formed exhibit a novel causal power: the power to generate phenomenal states, which they themselves instantiate: states that "interpret" what is going on in the brain.

5. Conclusion.

The purpose of this chapter has been to locate phenomenal states within a systematic metaphysical picture of the world, against two background assumptions: (1) phenomenal properties are fundamental, and (2) P-I properties, and the ways that they can bind together, are to be included in the class of fundamental phenomenal properties. I proceeded by beginning with the most ontologically conservative option and proceeded to wax inflationary as pressures to do so emerged. But in order to get a clear picture of the view that has emerged, it may be helpful to proceed in the opposite direction: if we begin with a picture on which mental phenomena are minimally dependent on physical phenomena, how far can we integrate the two types of phenomena while preserving what's unique about both?

---

this is the same sense in which metabolizing and processing sensory inputs and regulating my heartrate are things that I do.

Let us begin with the final proposal of the last section: conscious mental states are very much like brain-interpretations. Suppose we took this quite literally. How might it be possible? An interpreter (let's say God) looks at what's going on in the brain and ventures an interpretation, translates this translations this interpretation into the language of sensation and intentionality, and plops it into a single subject at a time. Maybe God has a series of switches, each corresponding to a phenomenal property. Some of the switches include modulators for number and intensity of instantiations as well as some mechanism for binding certain of the properties together. He looks at the brain at a time, sets the switches, and then outputs the resulting phenomenal state to the subject whose brain he is observing. This is a fanciful version of what was perhaps Leibniz's view.

Let's say that God wanted to cut himself out of the causal loop as much as possible. How would he do so? Essentially, he would need to build into the subject the capacity to instantiate suitable phenomenal properties, without an intervening interpreter. This would perhaps be Descartes' view. But Descartes held that the subject is a distinct thing from the body and not dependent upon it—hence God's causal activity remains crucial.

As a next step, God might make the subject—which are now understanding as the bearer of a brain-interpreting power—itself ontologically dependent on the brain, either by being a nonphysical simple that does not overlap the brain at all, or by being a physical composite that overlaps some or all of the brain. (As we have seen, it has to be a pretty special physical composite.)

It's worth noting how God can bow out still further so long as fairly low-level nervous systems are sufficient to generate subjects. That is to say: if subjects have the

capacity (1) to instantiate phenomenal interpretations of the brains that generate them, and (2) *to have an effect on the behavioral outputs of the brain on the basis of these phenomenal interpretations*, then more primitive minds will be able to develop into more sophisticated minds, across both ontogenetic development and phylogenetic development. That is, if subjects are not only brain-readers but also brain-programmers, then minds can grow, and mentally-endowed species can evolve.

This is as far as God can go, in the task of merging the mental and the physical. We are thus left with a version of interactionist dualism. Some will consider interactionist dualism an unacceptable theoretical stopping-point, because it commits us to a disunified picture of the world. But not all interactionist dualisms are on the same footing in this respect. I offer the foregoing as a comparatively integrated and elegant version.

<center>References</center>

Aizawa, Kenneth & Gillett, Carl. (2009). "Levels, Individual Variation and Massive Multiple Realization in Neurobiology." In John Bickle (ed.), *The Oxford Handbook of Philosophy and Neuroscience*. Oxford University Press. 539--582.

Allen, Colin and Trestman, Michael. (2015). "Animal Consciousness", *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2015/entries/consciousness-animal/>.

Baars, Bernard & Gage, Nicole. (2010). *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*. Elsevier.

Barnes, Elisabeth. (2014). "Fundamental Indeterminacy." *Analytic Philosophy* 55: 339-62.

Barsalou, Lawrence. (2008). "Grounded Cognition." *Annual Review of Psychology* 59: 617-645.

Bayne, Tim. (2009). "Perception and the Reach of Phenomenal Content." *Philosophical Quarterly* 59: 385-404.

Bayne, Tim & Chalmers, David. (2003). "What is the Unity of Consciousness?" In Axel Cleeremans (ed.), *The Unity of Consciousness*. Oxford University Press.

Bennett, Karen. (ms). "Why I Am Not a Dualist."

Block, Ned. (1986). "Advertisement for a Semantics for Psychology". *Midwest Studies in Philosophy* 10: 615-78.

Bickhard, Mark. (2004). "The Dynamic Emergence of Representation". In Hugh Clapin (ed.), *Representation in Mind*. Elsevier. 71-90.

_____. (ms). "Interactivism: A Manifesto."

Boghossian, Paul. (1989). "Content and Self-Knowledge". *Philosophical Topics* 17: 5-26.

Booth, Amy & Waxman, Sandra. (2003). "Mapping Words To The World In Infancy:

    Infants' Expectations For Count Nouns And Adjectives." *Journal of Cognition*

    and Development, 4: 357-381.

Brogaard, Berit. (2011). "Are There Unconscious Perceptual Processes*?" Consciousness*

    and Cognition 20: 449-63.

_____. (2012). "Split Brains: Brain's Processing of Information Affected by Hemispheric

    Transfer." *The Superhuman Mind. Psychology Today.* Web.

Brueckner, Anthony & Ebbs, Gary. (2012). *Debating Self-Knowledge*. Cambridge

    University Press.

Buckner, Cameron. (2013). "A Property Cluster Theory of Cognition." *Philosophical*

    Psychology 3: 1-30.

Burge, Tyler. (1988). "Individualism and Self-Knowledge". *Journal of Philosophy* 85:

    649-63.

_____. (2005). "Disjunctivism and Perceptual Psychology". *Philosophical Topics* 33:1-

    78.

_____. (2010). *Origins of Objectivity*. Oxford University Press.

Campbell, John. (2002). *Reference and Consciousness*. Oxford University Press.

Carey, Susan. Susan Carey (1999). "Knowledge Acquisition: Enrichment or Conceptual

    Change?" In Eric Margolis & Stephen Laurence (eds.), *Concepts: Core Readings*.

    MIT Press. 459-487.

_____. (2009). *The Origin of Concepts*. Oxford University Press.

_____. (2011). "Précis of the Origin of Concepts." *Behavioral and Brain Sciences* 34:
113-124.

Carruthers & Veillet, Benedicte. (2012). "The Case Against Cognitive Phenomenology."
In Tim Bayne & Michelle Montague (eds.), *Cognitive Phenomenology*. Oxford
University Press. 35-56.

Chalmers, David. (1995). "Facing Up to the Problem of Consciousness*." Journal of
Consciousness Studies* 2: 200-219.

_____. (1996). *The Conscious Mind*. New York: Oxford UP.

_____. (2000). "What is a Neural Correlate of Consciousness?" In Thomas Metzinger
(ed.), *Neural Correlates of Consciousness*. MIT Press. 17-39.

_____. (2003). "Consciousness and its Place in Nature." In The Blackwell Guide to the
philosophy of Mind, ed. P. Stich and T. Warfield: 102-42. Oxford: Blackwell.

_____. (2004). "The Representational Character of Experience." In Brian Leiter (ed.),
The Future for Philosophy. Oxford University Press. 153-181.

_____. (ms). "The Combination Problem for Panpsychism."

Chalmers, David & Clark, Andy. (1998). "The Extended Mind." Analysis 58: 7-19.

Chalmers, David & Jackson, Frank. (2001). "Conceptual Analysis and Reductive
Explanation." *Philosophical Review* 110: 315-61.

Chisholm, Roderick. (1957). *Perceiving: A Philosophical Study*. Cornell University
Press.

Clark, Andy. (2001). *Mindware: An Introduction to the Philosophy of Cognitive Science*.
Oxford University Press.

Cohen, Jonathan. (2008). "Colour Constancy as Counterfactual." *Australasian Journal of Philosophy* 86:61-92.

_____. (2009). *The Red and the Real: An Essay on Color Ontology.* Oxford University Press.

Craddock, et al. (2014). "Anesthetics act in quantum channels in brain microtubules to prevent consciousness." *Current Topics in Medicinal Chemistry*, 15: 523-533.

Dretske, Fred. 1981. *Knowledge and the Flow of Information*. MIT Press.

_____. (1986). "Misrepresentation." In R. Bogdan (ed.), Belief: Form, Content, and Function. Oxford University Press. 17-36.

_____. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.

_____. (1990). "Reply to Reviewers." *Philosophy and Phenomenological Research* 50: 819-839.

_____. (1995). *Naturalizing the Mind*. MIT Press.

_____. (1996). "Phenomenal Externalism: Or, If Meanings Ain't in the Head, Where Are Qualia?" *Philosophical Issues* 7: 143-58.

Egan, Andy. (2006). "Secondary Qualities and Self-Location." *Philosophy and Phenomenological Research* 72: 97-119.

Farkas, Katalin. 2013). "Constructing a World for the Senses." In Kriegel, ed., *Phenomenal Intentionality*. New York: Oxford UP.

Fillmore, Charles. (1982). "Frame Semantics." In The Linguistic Society of Korea, ed. *Linguistics In The Morning Calm: International Conference On Linguistics*. Hanshin Pub. Co. 111–139.

Fodor, Jerry. (1974). "Special Sciences." *Synthese* 28 (2):97-115.

_____. (1975). "The Language of Thought." Harvard University Press.

_____. (1987). "Psychosemantics: The Problem of Meaning in the Philosophy of Mind." MIT Press.

_____. (1990). *A Theory of Content and Other Essays*. MIT Press.

_____. (1994). *The Elm and the Expert*. MIT Press.

_____. (1999). *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press.

_____. (2008). LOT 2: The Language of Thought Revisited. Oxford University Press.

Fodor, Jerry & Lepore, Ernest. 1991. "Why Meaning (Probably) Isn't Conceptual Role." *Mind and Language* 6: 328-43.

_____. (1994). "What is the Connection Principle?" *Philosophy and Phenomenological Research* 54: 837-45.

Fodor, Jerry & Pylyshyn, Zenon. (1988). "Connectionism and Cognitive Architecture." *Cognition* 28: 3-71.

Frege, Gottlob. (1890/1997). "On Concept and Object." In Michael Beaney, (ed.). *The Frege Reader*. 181-193

Gazzaniga, Michael. (2002). "The Split Brain Revisited." *Scientific American*. 27-31.

Gelman, Rochel. (2011). "The Case for Continuity." *Behavioral and Brain Sciences* 34: 127-128.

Gentner, Dedre. (2010). "Bootstrapping the Mind: Analogical Processes and Symbol Systems." *Cognitive Science* 34: 752-775.

Gertler, Brie. (2007). Overextending the Mind? In Brie Gertler & Lawrence Shapiro (eds.), *Arguing About the Mind*. Routledge. 192-206.

Goff, Philip. (2009). "Why Panpsychism Doesn't Help Us Explain Consciousness."

    Dialectica 63: 289-311.

_____. (forthcoming.) "The Phenomenal Bonding Solution to the Combination Problem."

    In L. Jaskolla (ed.), *Panpsychism*. Oxford University Press.

Gopnik, Alison. (1997). "The Scientist as Child." *Philosophy of Science* 63: 485-514.

_____. (2001). "A Unified Account Of Abstract Structure And Conceptual Change:

    Probabilistic Models And Early Learning Mechanisms." *Behavioral and Brain*

    Sciences 34: 129-130.

Hameroff, Stuart & Penrose, Roger. (2014). Consciousness in the Universe: A Review of

    the 'Orch OR' Theory." *Physical Life Review* 11: 39-78.

Hanks, Peter. (2011). "Structured Propositions as Types." *Mind* 120: 11-52.

Harman, Gilbert. 1982. "Conceptual Role Semantics". *Notre Dame Journal of Formal*

    Logic 28: 242-56.

_____. (1990). "The Intrinsic Quality of Experience." *Philosophical Perspectives* 4: 31-

    52.

Hasker, William. (1999). *The Emergent Self.* Cornell University Press.

_____. (ms). "Do My Quarks Enjoy Beethoven?"

Heil, John. 1988. "Privileged Access". *Mind* 98: 238-51.

Hill. ms. "How Concepts Hook Onto the World".

Horgan, Terence. (2012). "Introspection About Phenomenal Consciousness: Running the

    Gamut From Infallibility to Impotence." In Declan Smithies & Daniel Stoljar

    (eds.), *Introspection and Consciousness*. Oxford University Press. 129-168.

Horgan, Terence & Tienson, John. (2001). "Deconstructing New Wave Materialism." In *Physicalism and Its Discontents*, ed. C. Gillett & B. Loewer: 307-18. Cambridge: Cambridge UP.

_____. (2002). "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press. 520-533.

Horgan, Terence & Graham, George. 2012. "Phenomenal Intentionality and Content Determinacy." In Richard Schantz (ed.), *Prospects for Meaning*. De Gruyter.

Horgan, Terence, Tienson, John & Graham, George. (2004). "Phenomenal Intentionality and the Brain in a Vat." In Richard Schantz (ed.), *The Externalist Challenge*. Walter De Gruyter.

Howell, Robert. 2006. "Self-Knowledge and Self-Reference." *Philosophy and Phenomenological Research* 72: 44-70.

Jackendoff, Ray. (1977). *X-Bar Syntax: A Study of Phrase Structure*. MIT Press.

_____. (1987). *Consciousness and the Computational Mind*. MIT Press.

Jackson, Frank. 1975. "On the Adverbial Analysis of Visual Experience." *Metaphilosophy* 6: 127-135.

_____. (1982). "Epiphenomenal Qualia." *Philosophical Quarterly* 32:127-36.

Jespersen (2012). "Recent Work on Structured Meaning and Propositional Unity." *Philosophy Compass* 7: 620-630.

Johnston, Mark. 1992. "How to Speak of the Colors." *Philosophical Studies* 68: 221-263.

Kelso, Scott. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. MIT Press.

Kinzler, Katherine & Spelke, Elizabeth. (2007). "Core Systems in Human Cognition." *Progress in Brain Research* 164: 257-264.

Kiss, Szabolcs. (2011). "The Notion Of Incommensurability Can Be Extended To The Child's Developing Theories Of Mind As Well." *Behavioral and Brain Sciences* 34: 134-135.

Kriegel, Uriah. (2011). *The Sources of Intentionality*. Oxford University Press.

_____. (2015). *The Varieties of Consciousness*. Oxford University Press.

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

Lakoff, George. (1999). "Cognitive Models and Prototype Theory." In Eric Margolis & Stephen Laurence (eds.), *Concepts: Core Readings*. MIT Press. 355-390.

Lakoff, George & Johnson, Mark. (1980). *Metaphors We Live By*. University of Chicago Press.

Langacker, Ronald. (1991). *Foundations of Cognitive Grammar, Vol. 2: Descriptive Applications*. Stanford University Press.

LaRock, Eric. (2007). "Disambiguation, Binding, and the Unity of Visual Consciousness." *Theory and Psychology* 17: 747-77.

_____. (2013) "From Biological Naturalism to Emergent Subject Dualism." *Philosophia Christi* 15: 97-118.

Laurence, Stephen & Margolis, Eric. (1999). "Concepts and Cognitive Science." In Eric Margolis & Stephen Laurence (eds.), *Concepts: Core Readings*. MIT. 3-81.

_____. (2012). "Abstraction and the Origin of General Ideas*." Philosophers' Imprint* 12: 1-22.

Levine, Joseph. (1983). "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354-61.

_____. (2007). "Phenomenal Concepts and the Materialist Constraint." In T. Alter & S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge*. Oxford UP. 145-66.

Lewis, David. (1983). "New Work for a Theory of Universals." *The Australasian Journal of Philosophy*, 61: 343-77.

Liberman, Alvin & Mattingly, Ignatius. (1985). "The Motor Theory of Speech Perception Revised." *Cognition* 21: 1-36.

Llinás, Rodolfo. (2003). "Consciousness and the Thalamocortical Loop." *International Congress Series*, 1250: 409-416.

Loar, Brian. (1990). "Phenomenal States." *Philosophical Perspectives* 4:81-108.

_____. (2003a). "Transparent Experience and the Availability of Qualia." In Quentin Smith & Aleksandar Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.

_____. (2003b). "Phenomenal Intentionality as the Basis of Mental Content." In Martin Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. MIT Press. 229-258.

Lormand, Eric. (1996). "Nonphenomenal Consciousness." *Noûs* 30: 242-61.

Ludwig, Kirk. (1996). "Explaining Why Things Look the Way They Do." In Kathleen Akins (ed.), *Perception, Vancouver Studies in Cognitive Science*, v. 5. Oxford UP: 18-60.

_____. (2012). "What Role Should Propositions Have in the Theory of Meaning? Review Essay: Scott Soames, *What is Meaning*?" *Philosophia*, 40: 885-901.

_____. (2013). "Methods in Analytic Epistemology." In Matthew Haug, ed. *Philosophical Methodology: The Armchair or the Laboratory?*. Routledge. 217-240.

Lycan, William. (2008). "Phenomenal Intentionalities." *American Philosophical Quarterly* 45: 233 - 252.

Margolis, Eric & Laurence, Stephen. (2013). "In Defense of Nativism." *Philosophical Studies* 165: 693-718.

Marr, David. (1982). *Vision.* Freeman.

Masrour, F. (2013). "Phenomenal Objectivity and Phenomenal Intentionality: In Defense of a Kantian Account." In Kriegel, ed., *Phenomenal Intentionality*. New York: Oxford UP.

McGinn, Colin. (1988). "Consciouness and Content." *Proceedings of the British Academy* 74: 225-245.

_____. (2005). *Mindsight: Image, Dream, Meaning*. Harvard University Press.

McKinsey, Michael (1991). "Anti-Individualism and Privileged Access". *Analysis* 51: 9-16.

Medin, Douglas. L. & Ortony, Andrew. (1989). "Psychological essentialism." In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. New York: Cambridge University Press.

Merker, Bjorn. (2005). "The Liabilities of Mobility: A Selection Pressure for the Transition to Consciousness in Animal Evolution." *Conscious and Cognition* 14: 89–114.

Millikan, Ruth. G. (1984). *Language, Thought and Other Biological Categories*. MIT Press.

_____. (1990). "Truth, Rules, Hoverflies, and the Kripke-Wittgenstein Paradox." *Philosophical Review* 99: 323-53.

_____. (1993). "Knowing What I'm Thinking Of". *Proceedings of the Aristotelian Society* 67: 91-108.

_____. (2009). "Biosemantics". In Brian P. McLaughlin & Ansgar Beckerman (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford University Press. 281-97.

Montague, Michelle. (2012). "The Phenomenology of Particularity." In T. Bayne & M. Montague (eds) *Cognitive Phenomenology*. Oxford UP. 121-140.

_____. (2013). "The Access Problem." In U. Kriegel (ed) *Phenomenal Intentionality: New Essays*. Oxford UP. 27-48.

_____. (2015). "Cognitive Phenomenology and Conscious Thought." Forthcoming in *Phenomenology and the Cognitive Sciences*.

Nagel, Thomas. (1979). "Panpsychism." In *Mortal Questions*. Cambridge University Press. 181-195.

Nes, Anders. (forthcoming). "The Sense of Natural Meaning in Conscious Inference." In T. Breyer & C. Gutland (eds.), *Phenomenology of Thinking*. Routledge.

Nida-Rümelin, Martine. (2007). "Grasping Phenomenal Properties." In *Phenomenal Concepts and Phenomenal Knowledge*, ed. T. Alter & S. Walter: 210-48. New York: Oxford UP.

_____. (2007). "Doings and Subject Causation." *Erkenntnis* 67: 255 - 272.

_____. (2011). "Phenomenal Presence and Perceptual Awareness: A Subjectivist Account of Perceptual Openness to the World". *Philosophical Issues* 21: 352-83.

O'Callaghan, Casey. (2010). "Experiencing Speech." *Philosophical Issues* 20: 305–332.

_____. (2011.) "Against Hearing Meanings." *The Philosophical Quarterly* 61: 783–807.

O'Connor, Timothy & Jacobs, Jonathan. (2003). "Emergent Individuals." *Philosophical Quarterly* 53: 540-555.

_____. (2010). "Emergent Individuals and the Resurrection." *European Journal for Philosophy of Religion* 2.

O'Connor, Timothy & Wong, Hong Yu. (2005). "The Metaphysics of Emergence." *Noûs* 39: 658-678.

O'Donahue, John. (1998). *Anam Cara: A Book of Celtic Wisdom.* Harper Perennial.

Papineau, David. (2002). *Thinking About Consciousness*.

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

_____. (1996). "Entitlement, Self-Knowledge, and Conceptual Redeployment". *Proceedings of the Aristotelian Society* 96:117-58.

_____. (1998). "Conscious Attitudes, Attention, and Self-Knowledge." In C. Wright, B. Smith & C. Macdonald (eds.), *Knowing Our Own Minds*. Oxford University Press. 63-98.

_____. (1999). *Being Known*. Oxford University Press.

Pitt, David. (2004). "The Phenomenology of Cognition (Or, What is it Like to Think that P?)"

_____. (2009). "Intentional Psychologism." *Philosophical Studies* 146: 117 - 138.

_____. (2011). "Introspection, Phenomenality, and the Availability of Intentional Content." In Tim Bayne & Michelle Montague (eds.), *Cognitive Phenomenology*. OUP. 141-173.

Prinz, Jesse. (2011). "The Sensory Basis of Cognitive Phenomenology." In Tim Bayne & Michelle Montague (eds.), *Cognitive Phenomenology*. OUP. 174-196.

Prior, E., Pargetter, R. & Jackson, F., 1982, 'Three Theses about Dispositions', *American Philosophical Quarterly* 19: 251–257.

Putnam, Hilary. (1967). "Psychological Predicates." In W. H. Capitan & D. D. Merrill (eds.), *Art, Mind, and Religion*. University of Pittsburgh Press. 37-48.

Quine, Willard van Orman. (1960). *Word and Object*. MIT Press.

_____. (1961/1953). "On What There Is." In *From a Logical Point of View*. Harvard University Press. 1-19.

Roelofs, Luke. (2014). "Phenomenal Blending and the Palette Problem." *Thought: A Journal of Philosophy* 3: 59-70.

Rosch, Eleanore. (1999). "Principles of Categorization." In Eric Margolis & Stephen Laurence (eds.), *Concepts: Core Readings*. MIT Press. 189-206.

Russell, Bertrand. (1910). "Knowledge by Acquaintance and Knowledge by Description." *Proceedings of the Aristotelian Society* 11: 108--28.

Scholl, Brian. (2001). "Objects and Attention: The State of the Art." *Cognition* 80: 1-46.

Schaffer, Jonathan. (2003). "Overdetermining Causes." *Philosophical Studies* 114: 23-45.

Schwitzgebel, Eric. (2011). *Perplexities of Consciousness*. MIT Press.

Searle, John. (1980). "Minds, Brains and Programs," *Behavioral and Brain Sciences* 3: 417–57

_____. (1991). "Consciousness, Unconsciousness and Intentionality." *Philosophical Issues* 1: 45-66.

_____. (1992). *The Rediscovery of the Mind.* MIT Press.

Shagrir (2001). "Content, Computation and Externalism." *Mind* 110: 369-400.

Shani, Itay. (2008). "Against Consciousness Chauvinism." *The Monist* 91: 294-323.

Siegel, Susanna. (2010). *The Contents of Visual Experience*. Oxford University Press.

Siewert, Charles. (1998). *The Significance of Conscousness*. Princeton University Press.

Siewert, Charles. (2012). "On the Phenomenology of Introspection". In Declan Smithies & Daniel Stoljar (eds.), *Introspection and Consciousness*. Oxford University Press. 129-168.

Sklar, Asael, et al. (2012) "Reading and Doing Arithmetic Nonconsciously." *Proceedings of the National Academy of Sciences* 109: 19614-19619.

Smart, J. J. C. (1959). "Sensations and Brain Processes." *Philosophical Review* 68:141-56.

Smithies, Declan. (2011). "What is the Role of Consciousness in Demonstrative Thought?" *Journal of Philosophy* 108: 5-34.

_____. (2012). "The Mental Lives of Zombies." *Philosophical Perspectives* 26: 343-372.

Soames, Scott. (2010). *What is Meaning?* Princeton University Press.

Spelke, Elizabeth. (2001). "Quinian Bootstrapping Or Fodorian Combination? Core And
Constructed Knowledge Of Number." *Behavioral and Brain Sciences* 34: 149-150.

Spelke, Elizabeth & Tsivkin, S. (2001). "Language and Number: A Bilingual Training
Study." *Cognition* 78: 45-88.

Stoljar, Daniel. (2005). "Physicalism and Phenomenal Concepts." Mind and *Language*
20: 296-302.

_____. (2009). "The Argument From Revelation". In Nola & Braddon-Mitchell (eds.),
Conceptual Analysis and Philosophical Naturalism. Cambridge: MIT Press.

Strawson, Galen. (1994). *Mental Reality*. MIT Press.

_____. (2008). *Real Materialism and Other Essays.* Oxford.

Tononi, Giulio and Koch, Christof. (2008). "The Neural Correlates of Consciousness: An
Update." *Annals of the New York Academy of Sciences*, 1124: 239–261.

Trestman, Michael. (2013a). The Cambrian Explosion and the Origins of Embodied
Cognition. *Biological Theory* 8: 80–92.

_____. (2013b). The Modal Breadth of Consciousness. *Philosophical Psychology* 6:1-19.

Tye, Michael. (1984). "The Adverbial Approach to Visual Experience." *Philosophical
Review* 93: 195-226.

_____. (1995). *Ten Problems of Consciousness*.

_____. (1999). "Phenomenal Consciousness: The Explanatory Gap as a Cognitive
Illusion." *Mind* 108: 705-25.

_____. (2009). *Consciousness Revisited*. MIT Press.

Tye, Michael & Briggs Wright. (2011). "Is There a Phenomenology of Thought?" In Tim Bayne & Michelle Montague (eds.), *Cognitive Phenomenology*. Oxford University Press. 326-344

Waxman, Sandra & Booth, Amy. (2001). "Seeing Pink Elephants: Fourteen-Month-Olds' Interpretations Of Novel Nouns And Adjectives." *Cognitive Psychology*, 43: 217-242.

Weizkrantz, et al (1974). "Visual Capacity In The Hemianopic Field Following A Restricted Occipital Ablation." *Brain* 97: 709-728.

White, S. (2010). "The Property Dualism Argument." In Koons & Bealer (eds.), *The Waning of Materialism: New Essays*. New York: Oxford UP.

Wilson, Jessica. (2013). "A Determinable-Based Account of Metaphysical Indeterminacy." *Inquiry* 56: 359-385.

Woodward, Philip. (ms). A Posteriori Physicalism and the Discrimination of Properties.

Zeki, Semir. (1993). *A Vision of the Brain*. Blackwell.

_____. (2003). "The Disunity of Consciousness." *Trends in Cognitive Sciences* 7: 214-218.

Zimmerman, Dean. (2010). "From Property Dualism to Substance Dualism." *Aristotelian Society Supplementary Volume* 84: 119 - 150.

Philip Woodward

Curriculum Vitae

Department of Philosophy, Indiana University
ppwoodwa@indiana.edu
www.philwoodward.org
Sycamore Hall 026, 1033 E. Third St.
Bloomington, IN 47405
(206) 228-3961

Areas of Specialization

Philosophy of Mind, Metaphysics

Areas of Competence

Philosophy of Science, Philosophy of Religion, Epistemology, Early Modern Philosophy

Education

Indiana University – PhD, 2015
     Thesis: *The Emergence of Mental Content: An Essay in the Metaphysics of Mind*
     Committee: Timothy O'Connor (chair), Colin Allen, Kirk Ludwig, Fred Schmitt

Indiana University – MA, 2009
     Thesis: "Constraining the Phenomenal Concept Strategy"
     Committee: Timothy O'Connor (chair), Fred Schmitt

Seattle Pacific University – BA, magna cum laude, 2005; Major in Philosophy

Dissertation Abstract: According to a common paradigm in philosophy of mind, for mental states to have intentional content is for those states to be suitably functionally embedded in a cognitive system vis-à-vis its environment. Such views, I argue, are unable to account for our first-personal knowledge of the content of our mental states. I develop an alternative approach according to which all intentionality is, at bottom, a way of being conscious. I call the relevant ways of being conscious "phenomenal-intentional properties." Along with the rest of the phenomenal realm, phenomenal-intentional properties emerge as ontologically novel properties of subjects. In perception, imagination and cognition, phenomenal-intentional properties combine with each other

(and with other non-intentional features of consciousness) to form arbitrarily complex and diverse representational states.

Publications

"Incarnation and the Multiverse," with Timothy O'Connor, forthcoming in Klaas Kraay ed., *God and the Multiverse*.

Book Review: *Introspection and Consciousness*, ed. Smithies & Stoljar, forthcoming in *Philosophical Psychology*.

In Preparation

"A Posteriori Physicalism and the Discrimination of Properties."

"The Acquaintance Argument for Intrinsic Intentionality."

"Phenomenal Intentionality: Reductionism vs. Primitivism."

"Consciousness and Cognitive Individuation."

"Three Choice-Points for Phenomenal Anti-Physicalism."

"The Semantic Structure of Phenomenal-Intentional States."

Presentations [*refereed]

"Consciousness and Cognitive Individuation," Consciousness Conference, University of Michigan-Flint, October 2015 *

"Emergent Materialism: An Inflationary Defense," Toward a Science of Consciousness, University of Helsinki, June 2015 *

"The Acquaintance Argument for Intrinsic Intentionality," Toronto Graduate Philosophy Conference, University of Toronto, May 2015 *

"Consciousness and Cognitive Individuation," Indiana Philosophical Association, Owensboro Community and Technical College, April 2015 *

"Phenomenal Intentionality: Reductionism vs. Primitivism," Syracuse Graduate Philosophy Conference, Syracuse University, March 2015 *

"Phenomenal Intentionality: Reductionism vs. Primitivism," Southeast Graduate Philosophy Conference, University of Florida, March 2015 *

"Phenomenal Intentionality: Reductionism vs. Primitivism," Society for Philosophy & Psychology, University of British Columbia, June 2014 *

"Phenomenal Intentionality: Reductionism vs. Primitivism," Indiana University, April 2014

"The Acquaintance Argument for Intrinsic Intentionality," Midsouth Philosophy Conference, Rhodes College, February 2014 *

"Mental Content and Transparent Access: Or, Why All Relational Theories of Intentionality Are Wrong," University of Evansville, April 2013

"A Posteriori Physicalism and the Principle of Rational Discrimination," Midsouth Philosophy Conference, Rhodes College, February 2013 *

"Trans-Universe Identity," with Timothy O'Connor, Midwest Annual Workshop in Metaphysics, St. Louis University, October 2012

"Emergence and Mental Causation," Undergraduate Research Symposium, University of Washington, May 2005

Honors/Awards

Graduate Essay Prize, 2015 Meeting of the Indiana Philosophical Association
James B. Nelson Dissertation Fellowship, 2013-2014

Teaching

Courses Designed & Taught:
- Introduction to Philosophy (fall 2011, fall 2012, spring 2013)
- Thinking and Reasoning (summer 2012)

Courses Taught As Assistant Instructor:
- Introduction to Ethics (with Andrew McAninch, fall 2008)
- Introduction to Existentialism (with Luke Phillips, spring 2012)
- Introduction to Philosophy (with Fred Schmitt, fall 2007; with Andrew McAninch, spring 2008; with Adam Leite, spring 2009)
- Philosophical Reflections on Religion & Evolution (with Timothy O'Connor, fall 2010)
- Public Oral Communication (with John Arthos, fall 2014)
- Thinking and Reasoning (with Jonathan Weinberg, spring 2011)

Undergraduate Courses Prepared to Teach:
(Lower-Division)
- Introduction to Ethics
- Introduction to Existentialism
- Introduction to Philosophy
- Critical Thinking
- Logic

(Upper-Division)
- Ancient Philosophy
- Early Modern Philosophy
- Epistemology
- Human Nature
- Metaphysics
- Philosophy of Mind
- Philosophy of Religion
- Philosophy of Science

Graduate Courses Prepared to Teach:
- Causality
- Intentionality
- Kinds, Concepts & Universals
- Objects & Properties
- Perception
- Philosophy of Mind
- Self-Knowledge
- The Given

Graduate Courses Taken

- 18th Century British Ethics (Kate Abramson)
- 20th Century Metaethics (Kevin Toh)
- Aristotelian Natural Philosophy (Carl Pearson)
- Causality (Jordi Cat)
- Kant's First Critique (Allen Wood)
- Logical Theory I (Joan Weiner)
- Modern Philosophy of Science (Jordi Cat)
- Objects in Space and Time (Timothy O'Connor)
- Phenomenology & Existentialism: Kierkegaard (Paul Spade)
- Philosophical Foundations of Cognitive Science (Jonathan Weinberg)
- Philosophy of Language: Analyticity (Gary Ebbs)
- Philosophy of Mind (Jonathan Weinberg)
- Plato: The Later Dialogues (Michael Morgan)

- Survey of the Philosophy of Science (Amit Hagar)

Extended Dissertation Abstract

All of us are aware that our perceptual, imaginative and cognitive states can be *directed at* or *about* things, and we regularly treat others as having such states as well. Philosophers have dubbed the phenomenon "intentionality" and have puzzled over several aspects of it: for one thing, if "aboutness" is a relation, one of its relata can be non-existent, even necessarily non-existent; for another, it is not clear how to locate the "aboutness" relation within the structure of the natural world. On my view, the reason such aspects of intentionality have seemed so mysterious is that philosophers have failed to appreciate the close relationship between intentionality and consciousness. My dissertation develops and defends a consciousness-based theory of intentional mental content.

Chapter 1: The Acquaintance Argument for Intrinsic Intentionality. One strategy for giving an account of intentional aboutness has been to locate it within the domain of the *causal*. Perhaps, that is, the aboutness-relation just is a kind of causal relation in reverse (or a disposition to respond in certain ways to causal relations of certain sorts); or perhaps causal relations somehow constitute *inferential* relations, and a thing's intentional content is fixed by its inferential profile. Such suggestions, often lumped together under the heading of "naturalized intentionality", have been dominant among physicalist philosophers of mind for the last thirty years or so. I argue that these strategies cannot supply a metaphysics of intentional directedness, because the properties they point to lack a key feature of some intentional properties: viz. the feature of being *intrinsic to* the states that include them. Were no intentional properties intrinsic to the states that instantiate them, we would be unable to recognize mental states of distinct modes (i.e. visual vs. auditory, or perceptual vs. cognitive) as having intentional contents in common.

Chapter 2: Motivating The Phenomenal Grounding Thesis. If intentional directedness is not a matter of dispositions to cause and be caused in certain ways, what is it? Along with a number of recent philosophers I maintain that intentional directedness is fundamentally *an aspect of consciousness*. In order to motivate this view (which I call "The Phenomenal Grounding Thesis"), I discuss several arguments in the literature according to which certain forms of consciousness are (a) *sufficient* for intentionality; and (b) *necessary* for intentionality.

Chapter 3: A Theory of Phenomenal Grounding. Though there is substantial literature *defending* the Phenomenal Grounding Thesis, not much has been said by way of its *explanation*—a task I take up in the remainder of the dissertation. *Phenomenal properties* are experiential aspects of consciousness. There are many types of phenomenal property, e.g. sensory, somatic and proprioceptive and conative properties. According to my theory, a subject's conscious mental states have the intentional content they do because they include phenomenal properties of a type distinct from all of those just mentioned: *phenomenal-intentional properties*. Example: the phenomenal-intentional property *being*

phenomenally-intentionally directed causality-wise, when instantiated by a subject, presents the property *being a cause* to that subject, and consequently *being a cause* is part of the intentional content of her conscious state. (A subject's *non-*conscious mental states are, on my view, her dispositions to instantiate conscious mental states under certain circumstances.).

*Chapter 4: The Semantic Structure of Phenomenal-Intentional States***.** Intentional states have a certain structure—often, though not always, a *propositional* structure. I show how such *semantic* structure can be explained in terms of the *metaphysical* structure of phenomenal-intentional states.

Chapter 5: Constructing and Abstracting the Intentional Repertoire. Mature human subjects can entertain intentional states of unbounded quantity and variety. I contend that a few primitive phenomenal-intentional properties (as of, e.g. objecthood, causality, spatial relations, agency, etc.) when properly embedded in perceptual, imaginative and cognitive states, can be used to construct all of these contents—or can be used, by the subject, as a way of leveraging her way to the instantiation of new primitive phenomenal-intentional properties as of all of these contents.

Chapter 6: The Emergence of Phenomenal Intentionality. How are phenomenal-intentional properties to be located within the natural order? On my view they amount to (along with the rest of consciousness) novel additions to reality, irreducible to physical systems on which they depend. Their emergence occurs according to a two-stage process: first, *subjects* emerge from physical systems, when those systems exhibit the right sort of organic unity. I explain such emergence in terms of the activation of latent capacities of the system's ultimate physical constituents. Second, subjects instantiate richly structured phenomenal-intentional states, in response to the particular functional arrangement at a time of such states' emergence bases. These phenomenal-intentional states are in turn causally relevant to the ongoing dynamic evolution of the emergence base. Put picturesquely, emergent subjects are—though are of course not aware that they are—*brain-interpreters* and *brain-programmers.*

Professional References

Timothy O'Connor
Professor of Philosophy
Indiana University
(812) 855-6817
toconnor@indiana.edu

Kirk Ludwig
Professor of Philosophy
Indiana University
(812) 855-2404
ludwig@indiana.edu

Fred Schmitt
Professor of Philosophy
Indiana University
(812) 855-3296
fschmitt@indiana.edu

Jonathan Weinberg
Associate Professor of Philosophy
The University of Arizona
(520) 621-9401
jmweinberg@email.arizona.edu