

The Varieties of Consciousness

URIAH KRIEDEL

OXFORD
UNIVERSITY PRESS

CONTENTS

Preface vii

Introduction: Phenomenal Primitives 1

1. Cognitive Phenomenology 38

2. Conative Phenomenology 72

3. The Phenomenology of Entertaining 97

4. Emotional Phenomenology 129

5. Moral Phenomenology 159

Conclusion: The Structure of the Phenomenal Realm 184

Appendix: Theses on the Phenomenology of Freedom 205

Notes 245

Bibliography 273

Index 283

Introduction

Phenomenal Primitives

Recent work on consciousness has featured a number of debates on the existence and character of controversial types of phenomenology. Perhaps the best-known is the debate over the existence of a *sui generis*, irreducible *cognitive phenomenology*—a phenomenology proper to thought. Another concerns the existence of a *sui generis* phenomenology of agency. Such debates bring up a more general question: how many types of *sui generis*, irreducible, basic, primitive phenomenology do we have to posit to just be able to *describe* the stream of consciousness? This book attempts to make some contributions toward answering this question. The purpose of this introduction is to clarify the question and how one might approach it, laying relevant metaphysical (§§2–5) and methodological (§§6–8) foundations.

1. Before (and after) Philosophy

About a century ago, stream-of-consciousness literature started being promulgated by such writers as James Joyce, Marcel Proust, and Arthur Schnitzler. Instead of the traditional well-organized, cleaned-up, highly processed presentation of characters' thought and conduct, these writers attempted to describe inner life in a realistic, hence somewhat chaotic and confused, fashion. We might say they were seeking a more phenomenologically adequate rendering of inner life. The technique has quickly gained popularity and renown. It is a good question what is so compelling about it, but one immediate thought is that it offers insight into questions of the form "What is it like to be this kind of person, in this kind of situation?" Outside fiction, we have *direct* insight only into our own consciousness; this epistemic loneliness is broken by stream-of-consciousness narratives that offer a believable peep into another consciousness.

Typical stretches of stream-of-consciousness prose often feature sensory perceptions, bodily sensations, and visceral aspects of emotional feelings. However, they are *never* restricted to such elements. On the contrary, they

also feature—at least as centrally—thought processes, hopes and desires, and more intellectual aspects of emotion. Without the latter, the resulting narratives would be extremely boring. The technique would have exhausted itself long ago. Nobody wants to read about an interminable sequence of sensory occurrences. But many of us want to read about stretches of inner life as described by stream-of-consciousness writers. This is because in describing also the thought processes, hopes and desires, and more intellectual aspects of emotion, these narratives manage to shed light on what it is like to occupy a different subjective perspective. Without including such elements, no real insight into what it is like to be someone else would be gained.

Consider an early example. In his 1897 short story “The Dead Are Silent,” Arthur Schnitzler explores a stretch of inner life of a married woman who flees the scene of an accident in which her lover just died. The accident occurs in a countryside corner just outside Vienna, and much of the story takes place “inside her head” as the protagonist makes her way back home to her husband and child. Here is Schnitzler’s description of her approach to Vienna’s city center:

The noise of the city grows louder, the street is lighter, the skyline of the Prater street rises before her, and she knows that she can sink into a flood tide of humanity there and lose herself in it. When she comes to a street lamp she is quite calm enough now to take out her watch and look at it. It is ten minutes to nine. She holds the watch to her ear—it is ticking merrily. And she thinks: “Here I am, alive, unharmed—and he—he—dead. It is Fate.” She feels as if all had been forgiven—as if she had never sinned. And what if Fate had willed otherwise? If it were she lying there in the ditch, and he who remained alive? He would not have run away—but then he is a man. She is only a woman, she has a husband, a child—it was her right—her duty—to save herself. She knows that it was not a sense of duty that impelled her to do it. But what she has done was right—she had done right instinctively—as all good people do. If she had stayed she would have been discovered by this time. The doctors would question her. And all the papers would report it next morning; she would have been ruined forever, and yet her ruin could not bring him back to life. Yes, that was the main point, her sacrifice would have been all in vain.

This passage starts with an enumeration of sensory elements: loud noises, brighter lights, and so on. Soon the action moves to the emotional domain, as a summary battle with a sense of guilt takes place; importantly, we do not grasp the phenomenology of the protagonist’s guilt through descriptions of her visceral sensations, but through descriptions of highly conceptual thought processes. Finally, a sort of emotional clarity dawns on the protagonist as a

purely cognitive, intellectual event takes place: she manages to articulate to herself something she was feeling when she *realizes* what matters in the situation, namely, that if she had stayed by her lover's side and sacrificed her reputation and her family's happiness, her sacrifice would have been in vain. Again, we would not really grasp what it is like to be her at that moment if we were told that her stream of consciousness included visual sensations of the shapes my^sacrifice^would^have^been^in^vain, or that it included auditory imagery of the corresponding sounds. It is the *meaning* of those shapes and sounds, the *proposition* present before her mind, that gives us the kind of insight that makes it worth reading stream-of-consciousness literature. And in any case, it is presumably not my^sacrifice^would^have^been^in^vain that floated before her mind's eye, but für^nichts^hätte^sie^sich^zu^Grunde^gerichtet. Yet readers who cannot read German still gain insight into the protagonist's inner life by being described this episode. The reason is that the correct description of the episode focuses not on any shape-sensations the protagonist may experience, but on the *thought*, as a contentful conscious event, that occurs to her.

Much of this book will be dedicated to philosophical arguments to the effect that there *exist* forms of nonsensory (phenomenal) consciousness, including cognitive and conative. I am alive to the possibility, however, that the opposite view may require therapy more than argumentation. My therapeutic prescription is a healthy daily dose of stream-of-consciousness literature. Just as Harman (1990) predicted that when you introspect seeing your hands you will become aware of nothing but the hands you are seeing, I issue the following prediction: the longer you immerse yourself in stream-of-consciousness literature, the odder would seem to you the notion that phenomenal consciousness involves no pure, irreducible cognitive and conative elements.

I. Metaphysical Foundations

2. Phenomenal Primitives and Phenomenal Grounding

According to eliminativists about phenomenal consciousness (e.g., Rey 1988), there are no phenomenal properties—at least no instantiated ones. Most philosophers of mind are not eliminativists here. They accept the existence of *some* types of phenomenology. In particular, there are two types of phenomenology traditionally thought relatively uncontroversial: the phenomenology of perceptual experience and the phenomenology of pleasure and pain (“algedonic” phenomenology). Mainstream philosophy of mind tends to accept those.

What about other types of phenomenology, neither perceptual nor algedonic? For example, what about a phenomenology of thinking, or of agency? One perfectly coherent and stable position is to deny the existence of any such, accepting *only* perceptual and algedonic phenomenology; this would be a sort

of eliminativism about the phenomenology of thought, of agency, and so on. On this view, there is nothing it is like to think, for example. Another option is to accept some further types of phenomenology, but claim that ultimately they result from combinations of perceptual and algedonic phenomenology; this would be a sort of reductivism rather than eliminativism. On this view, there is something it is like to think, but *what* it is like is just the right combination of sensory experiences. A third option, however, is to accept these more adventurous types of phenomenology and furthermore claim that they go beyond perceptual and algedonic phenomenology, constituting *sui generis* or *primitive* types of phenomenology; this would be a sort of *nonreductivism* or *primitivism*.

In fact, for each putative type of phenomenology, these three options are *prima facie* open. For each there is an eliminative, a reductive, and a primitivist option. Consider cognitive phenomenology. Some deny that there is any distinctive phenomenology associated with thinking that $2 + 2 = 4$ (Nelkin 1989 *inter alia*). Others accept the existence of a phenomenology of thinking that $2 + 2 = 4$, but argue that ultimately it amounts to some type of already familiar sensory phenomenology; often the idea is that it is the phenomenology of auditory imagery in “inner speech,” whereby the sounds “two plus two equal four” float by one’s mind’s ear, so to speak (Carruthers 2006 and Prinz 2011 *inter alia*). Still others, however, insist that there exists a *sui generis* cognitive phenomenology that goes beyond the phenomenology of inner speech, and in fact cannot be captured by any form of perceptual (or algedonic) phenomenology (Siewert 1998 ch.8 and Pitt 2004 *inter alia*). The first view is *eliminativist* about cognitive phenomenology, the second *reductivist*, and the third *primitivist*.

The choice among these three options can be appreciated through an inconsistent triad:

- 1) There exists cognitive phenomenology.
- 2) Cognitive phenomenology is irreducible to perceptual and algedonic phenomenology.
- 3) Perceptual and algedonic phenomenology ultimately exhaust all phenomenology.¹

The eliminativist about cognitive phenomenology denies 1, the reductivist denies 2, and the primitivist 3.

(It should be stressed that the debate over reductivism here is orthogonal to the debate over physicalism. The issue is not whether cognitive phenomenology reduces to *physical* properties, but whether it reduces to other *phenomenal* properties. One could be a primitivist about cognitive phenomenology in the sense of holding that it is irreducible to any other phenomenology and still reduce cognitive-phenomenal properties to some neural properties. Conversely, one could hold a reductivist account of cognitive phenomenology in terms of

perceptual phenomenology but be a dualist about phenomenal properties in general, thus denying the physical reducibility of cognitive-phenomenal properties. The two issues are logically independent. My concern here is with the *phenomenal* reducibility, not *physical* reducibility, of certain types of phenomenology.)

Similar choices face us with respect to other putative types of phenomenology. Consider emotional phenomenology. It is odd, but not incoherent, to maintain that there is nothing it is like to undergo emotional episodes; the eliminativist about phenomenal consciousness is presumably committed to this. More commonly, it has often been argued that the phenomenal feel of emotion is nothing but proprioceptive (or kinesthetic, or somatic) feeling (James 1884 and Armstrong 1968 *inter alia*). This is to claim that emotional phenomenology reduces to a species of perceptual phenomenology. At the same time, it is also possible to hold that whatever other elements emotional experiences involve, there is also a more basic *sui generis* element that constitutes a primitive emotional phenomenology (Stocker 1996 and Montague 2009 *inter alia*). The choice among these views can again be captured in a triad:

- 1) There exists emotional phenomenology.
- 2) Emotional phenomenology is irreducible to perceptual and/or algedonic phenomenology.
- 3) Perceptual and algedonic phenomenology ultimately exhaust all phenomenology.

Similar triads can be formulated for other types of phenomenology—the phenomenology of agency, for instance (see Bayne 2008).



It is worth noting that the three theoretical options do not *have* to be formulated relative to perceptual and algedonic phenomenology as potential reducers. A primitivist about cognitive phenomenology who considered the status of emotional phenomenology should include cognitive phenomenology among the potential reducers. At the same time, someone could deny that perceptual or algedonic phenomenology is primitive. For example, she might hold that pain is just tactile perception of tissue damage, effectively reducing algedonic to perceptual phenomenology. For someone who accepted cognitive phenomenology but not algedonic phenomenology as primitive, the question of emotional phenomenology would become this: does emotional phenomenology exist, and does it reduce to some combination of perceptual and cognitive phenomenology?²

The general issue for any putative phenomenology P, then, is whether it reduces to whatever primitive phenomenologies one already recognizes. Put in terms of an inconsistent triad, the general form of the question, for any given

putative phenomenology P , and phenomenologies P_1, \dots, P_n recognized as primitive, is what to reject among:

- 1) There exists phenomenology P .
- 2) P does not reduce to phenomenologies P_1, \dots, P_n .
- 3) P_1, \dots, P_n exhaust all phenomenology.

Call this the *generalized triad* for phenomenology.

With this generalized triad, we can answer straightforwardly the question “How many types of phenomenology do we need to posit to just be able to *describe* the stream of consciousness?” The answer is: as many types as primitivism (rather than reductivism or eliminativism) is true of. There is a view, which we may call “mainstream stingy-ism,” that is primitivist only about perceptual and algedonic phenomenology.³ My starting point in this book is that mainstream stingy-ism is likely false, and there probably exist some non-perceptual, non-algedonic types of irreducible phenomenology. The question is which ones they are.

The generalized triad can help us put in place an *idealized, or rationally reconstructed*, “procedure” for generating coherent and stable accounts of the scope of primitive phenomenology. To a *first approximation*, the “procedure” is this. In Step 1, we produce a comprehensive inventory of *putative* types of phenomenology. In Step 2, we feed each item in this inventory into the generalized triad just described, with that item serving as a substitution instance for P and all other items functioning as P_1, \dots, P_n ; going through each of these triads, we attempt to establish which values of P are such that the primitivist position is the most plausible for them. (*How* we do so is the topic of §§6–8.) This produces a first outcome: a list of all phenomenal primitives. That is, it divides the set of all putative phenomenologies into two subsets: the primitive ones and the rest. Moving now to Step 3, we feed each member of the non-primitive subset into a new triad in which P_1, \dots, P_n are given by all and only the members of the primitive subset; here we attempt to establish which values of P are such that reductivism is more plausible for them than eliminativism. This produces a second outcome: a list of all “phenomenal derivatives.” That is, it divides the non-primitive subset into two further subsets: those that are derivative upon the phenomenal primitives and those that are not.⁴ The overall outcome is a structure we might describe as $(S_1, (S_2, S_3))$, where S_1 = the set of all phenomenal primitives, S_2 = the set of all phenomenal derivatives, and S_3 = the set of *mere putatives*.⁵ This structures the phenomenal realm along an important dimension, presenting all phenomenal properties and all reduction or “grounding” relations among them. It tells us what the phenomenal primitives are and what other phenomenal properties they ground/reduce.⁶

(I will be using “grounding” as interchangeable with “reduction.” In truth, this is quite incorrect, as grounding is typically taken to be an in-virtue-of relation, whereas reduction is not.⁷ In a way, though, my real interest is in the *disjunction* of grounding and reduction.)



As noted, this is only a first approximation of the “procedure,” and I will offer refinements shortly. Already in this form, however, we can appreciate that the project at hand is a special case of a more general metaphysical project. According to Schaffer (2009), the central mandate of metaphysics is to tell us what grounds what. Although philosophers commonly profess to be eliminativists about this or that putative entity, often closer inspection reveals that their view is better classified as reductivist, claiming that the relevant putative entity is “nothing but” some other entity or collection of entities (and thus is not *fundamental*, and no *addition of being*). The true goal of metaphysics, according to Schaffer, is to identify the basic, ungrounded entities in terms of which all other entities can be accounted for—the ungrounded grounders of reality, if you will.⁸ At the same time, genuine eliminativism is a real *option*, and sometimes it is clearly adopted—as when the atheist goes eliminativist with respect to gods. Arguably, metaphysics always presents us with a choice between primitivist, reductivist, and eliminativist positions. The present project can be seen as a metaphysics of phenomenology.

In an ostensibly more epistemological vein, Chalmers (2012) proposes that a central goal of philosophical worldviews is to produce the minimal base of truths from which the totality of truths could be derived *a priori* by an ideal reasoner (see already Jackson 1998). This is, in Chalmers’ terms, a “scrutability project.” (When p can be derived from q *a priori* by an ideal reasoner, we say that p is “scrutable” from q .) I describe the vein as only *ostensibly* epistemological because on the “Australian view” of ontological reduction (Chalmers and Jackson 2001), a necessary condition on the ontological reduction of entity E_1 to entity E_2 is that all truths about E_1 be scrutable from truths about E_2 . Against this background, questions of phenomenal *reducibility* implicate questions of phenomenal *scrutability*: which phenomenal truths are scrutable from which, and which types of phenomenology are such that truths about them form the scrutability base of all phenomenal truths? So against the background of the Australian view, the present project can also be seen as a scrutability project restricted to the phenomenal realm.

It might be objected that certain features of the phenomenal realm make this kind of project, legitimate in other areas of inquiry, unviable for phenomenology. For example, it might be argued that the kind of project outlined above presupposes that the putative entities up for elimination, reduction, or recognition as primitive must be discrete and separate from each other; but that there

is a perfectly reasonable view, which we may call *phenomenal holism*, that denies this. In other words, the project under discussion is beholden to a rather naïve sort of phenomenal atomism.

It is clear, however, that the project is fully consistent with phenomenal holism. But to see why, we need a more precise formulation of phenomenal holism. Chudnoff (2013a: 562) offers the following: “All partial phenomenal states of a subject at a time metaphysically depend on the subject’s total phenomenal state at that time.” Consider the overall experience of a bite of peanut-butter and jelly sandwich. This involves as parts (i) a gustatory experience of peanut butter, (ii) a gustatory experience of jelly, (iii) associated olfactory experiences, (iv) tactile experiences of wet bread, perhaps (v) a motivational experience of wanting the next bite or (vi) an aesthetic experience of appreciating the sandwich’s taste, and much more. According to holism, the overall experience does not metaphysically depend on (i)–(vi), but on the contrary (i)–(vi) depend on it: (i)–(vi) exist in virtue of the overall experience existing, not the other way round. We can see now that phenomenal holism does not exclude distinguishing different parts of an overall experience, it just makes a claim about metaphysical priority or dependence between the parts and the whole. As long as we have well-defined parts on our hands, we can then ask whether primitivism, reductivism, or eliminativism is the right view of them. Thus, even if the aesthetic experience of the sandwich metaphysically depends on the overall experience, we can wonder whether it involves a *sui generis* phenomenology or not.⁹

The objector might insist that there is also a stronger kind of phenomenal holism that denies the possibility of distinguishing different parts within a dated overall experience. On this view, (i)–(vi) is an artificial, pragmatically driven decomposition of the bite experience. In and of itself, the experience is strictly *indivisible*. Compare: in the metaphysics of material objects, there is a distinction between two kinds of “monism.” According to “priority monism,” individual objects metaphysically depend on the overall universe, such that facts about this dog or that tree obtain in virtue of facts about the overall universe obtaining (Schaffer 2007, 2010). According to “existence monism,” by contrast, there are simply no individual objects other than the overall universe—the latter is the only material object (Horgan and Potrč 2008). The phenomenal holism just considered is analogous to priority monism, but there is also a version of phenomenal holism analogous to existence monism. On that version, there are no partial experiences, characterized by a discrete phenomenology about which we can debate the merits of primitivism, reductivism, and eliminativism.

However, this stronger phenomenal holism faces a dilemma. On the face of it, the overall experiences we have at a time normally seem *structured*, not homogeneous blobs. It is natural to accommodate this by adverting to parts: the structure of a dated overall experience is fixed by the properties of its parts (even if these parts metaphysically depend upon the overall experience). The strong

holist faces a dilemma: either (a) she attempts to accommodate structure in some other fashion, or (b) she denies the datum of structure. If (a), then presumably some other notion will be invoked (“*dimension* of experience?”), but then our project could be framed in terms of that notion. If (b), then the view is truly implausible—implausible enough that its rejection is no longer a meaningful liability on our project.¹⁰

3. Phenomenal Determinables and Determinates

The project does face an immediate complication, however. Consider the difference between the phenomenology of seeing yellow and the phenomenology of seeing blue. Both are species of visual phenomenology, which in turn is a species of perceptual phenomenology. The fact that there is quite a bit of variety within perceptual phenomenology might induce some to reconsider its status as phenomenally primitive. After all, there is a sense in which perceptual phenomenology is but the collection of all its possible species: seeing yellow, seeing blue, seeing circles, hearing trumpets, and so on. At the same time, there is also a sense that the internal variety in perceptual phenomenology is irrelevant to its status as primitive: the mere fact that it does not reduce to phenomenologies algedonic, cognitive, emotional, and so on guarantees the status. We are tempted to say that perceptual phenomenology is primitive because it is irreducible to any phenomenology *at the same level of generality*.

Unfortunately, this way of putting things is somewhat vague. The true moral, it seems to me, is that the notion of phenomenal primitiveness is relative—it must be relativized to those “levels of generality.” One way to think of this is as follows: the structure of the phenomenal realm must refer not only to grounding relations but also to “determinable/determinate” relations or “genus/species” relations. Two phenomenologists may agree on which emotional experiences instantiate which phenomenal properties, but disagree on whether emotional phenomenology should be taken to have twelve species or fourteen. This appears to be a disagreement about the structure of the phenomenal realm, but not about grounding.

If this is right, then the above procedure for generating accounts of phenomenal primitives and derivatives requires modification. Recall that Step 1 in the procedure drew an inventory of putative types of phenomenology, Step 2 identified the phenomenal primitives in it, and Step 3 identified the phenomenal derivatives. It would now seem that the procedure requires two major modifications. First, an intermediary step is needed between Steps 1 and 2, whose purpose is to figure out the determinable/determinate and genus/species relations among the items in the inventory. Secondly, when we feed items into triads in Steps 2 and 3, we should consider as potential reducers only phenomenologies from the same level.

Before expanding on these two modifications, a word on the relationship between the determinable/determinate and genus/species relations. Both relations can be found in the phenomenal realm: visual and auditory are two different species of perceptual phenomenology, but reddish₁₆ and reddish₁₇ are two different *determinates* of visual phenomenology. There are two main differences between them. First, there is a continuum between determinates of the same determinable (such as reddish₁₆ and reddish₁₇ qualities) but not between species of the same genus (such as auditory and visual qualities). Secondly, determinates of the same determinable necessarily exclude each other, whereas species of the same genus do not (an experience of a loud blue airplane overhead can be both visual and auditory but not both bluish₂₃ and bluish₂₄).¹¹ It would be useful to have a term for the generic relation two species of which are the determinate/determinable and genus/species relations: “encompassing/encompassed” might be an option.¹² Instead, however, I am going to use “determinable/determinate” technically to cover both relations.

The highest phenomenal determinable is *phenomenality per se* (what-it-is-like-ness as such, if you will).¹³ It is the phenomenal property that is not a determinate of any other phenomenal property. The second-highest phenomenal determinables are those phenomenal properties which are determinates of no other phenomenal property but *phenomenality per se*. They are determinates of only one phenomenal determinable. Third-layer phenomenal determinables are determinates only of *phenomenality per se* and of phenomenal properties that are determinates only of *phenomenality per se*. And so on and so forth. Presumably, there is also a bottom layer of *maximally determinate* phenomenal properties. These are phenomenal properties that do not serve as determinables of any other phenomenal properties. Brentano, Wundt, and the early introspectionists called these the *elements of consciousness*. One task, then, is to stratify the initial inventory of putative phenomenologies into the number of determinable layers between the highest determinable of *phenomenality per se* and the maximally determinate “elements” of consciousness.¹⁴ (There is a question as to how we might *establish* which phenomenal determinables a putative type of phenomenology is a determinate of; I will return to this in §4.)

This stratification alters the “procedure” from §2. Suppose we are concerned to establish whether some phenomenology *P* is primitive. This requires that the reducibility of *P* be considered within the right “layer” of phenomenal properties. More precisely, it requires that we (i) determine the *n*-layer *P* belongs to, (ii) identify in our inventory all the other *n*-layer putative phenomenologies P_1, \dots, P_n , and (iii) feed *P* into a triad in which P_1, \dots, P_n serve as *P*’s potential reducers.

A *second approximation* of the “procedure” would thus involve *four* steps. Step 1 creates the initial inventory of putative phenomenologies. Step 2 stratifies these into layers of determinate-ness. Step 3 identifies the phenomenal primitives in each layer. Step 4 identifies the phenomenal derivatives in each layer.

The overall outcome is a conjunction of lists each of which specifies phenomenal primitives, derivatives, and mere putatives in different layers of phenomenal determinate-ness.



Again certain worries about the viability of the envisaged project may arise. One is that, for all we know, there are several equally good ways to organize phenomenal properties into layers of determinacy, and there are no objective facts of the matter as to which one is the “real one.” This is certainly a fair worry, which will be taken up in §5.

Another possible worry is that talk of phenomenal determinables can be indulged as a *façon de parler*, but cannot be taken literally. In general, an objector might contend, there are no determinable properties at all. Although we have *concepts* for such properties, there is an immediate causal-preemption threat: determinable properties’ causal powers are presumably exhausted by their determinates’, rendering the former explanatorily redundant (Gillet and Rives 2005). Thus the only real properties are maximally determinate ones: in the phenomenal case, the aforementioned phenomenal “elements.”

There are three responses to this objection. First, the debate over the ontological status of determinables is by no means resolved. Wilson (2012) argues (roughly) that determinables have certain modal properties that cannot be accounted for by their determinates, and are therefore irreducible to them. Consider a red car that is in fact red_{17} . The car’s redness has the property that it could have been another shade of red (say, red_{21}). This modal property of the car’s redness is inexplicable in terms of the car’s red_{17} -ness. For it is not true that the car’s red_{17} -ness could have been a red_{21} -ness. To imagine that the car is red_{21} rather than red_{17} is to imagine that the car’s red_{17} -ness does not exist at all, not that it exists but with a different character or quality. But to imagine this is *not* to imagine that the car’s *redness* does not exist at all; rather, it is precisely to imagine that the car’s redness has a different quality.¹⁵ This is not the place to pursue a close examination of the cogency of Wilson’s modal argument. The argument is at least initially compelling and bears further consideration, especially since it parallels modal arguments about material constitution.¹⁶ Accordingly, it is far from settled that there are no determinables.

Secondly, *concepts* for phenomenal determinables are of great value even if it turns out that there are no *properties* they pick out. As natural, evolved creatures of limited cognitive resources, in grasping a realm of phenomena we cannot operate with concepts for maximally determinate properties only. Putting order in the phenomena at the level of determinacy at which we interact with them requires that we develop concepts for determinables. This is why we operate in everyday life mostly with the concept of red and not red_{17} . And while a scientific framework is supposed to refine the prescientific grain, no existing science does

away with concepts for determinables altogether. Thus insofar as the present goal is to develop a framework through which to study the phenomenal realm and put order in its phenomena, determinable-concepts are indispensable. The project could always be recast as concerning the structure of the conceptual scheme for making sense of the phenomenal realm (rather than as concerning the structure of the phenomenal realm itself).

Thirdly, the absence of phenomenal determinables would not vitiate the part of the project concerned with grounding relations. We could ask whether there is *sui generis* cognitive or conative phenomenology and mean by that no more than whether there are irreducible maximally determinate cognitive-phenomenal or conative-phenomenal properties.

In conclusion, assessment of the status of putative phenomenologies as primitive, derivative, or *merely* putative must be conducted relatively to a layer of determinate-ness. This book purports to make first steps on one sliver of this overall project. The contribution envisaged is to the following question:

- (Q) Which putative types of phenomenology are primitive second-layer phenomenal determinables?

The bulk of this book addresses Q by considering five putative types of phenomenology.

4. Putative Phenomenal Universals

We have discussed two dimensions of structure in the phenomenal realm: grounding relations and determinable/determinate relations. There are surely many others. Some phenomenal properties may bear one-way dependence relations to other phenomenal properties. Other property pairs may bear two-way dependence relations. Some phenomenal properties bear structural resemblance relations to some but not all other phenomenal properties. There is thus a group of highly general relations, which we may call *metaphysical relations*, that defines the *overall structure of the phenomenal realm*.¹⁷

In this book, my concern is only with those dimensions of the overall structure of the phenomenal realm that bear on Q. This involves in the first place grounding and determinable/determinate relations. To fully understand these dimensions of phenomenal structure, however, we must also understand the *relata* bearing grounding and determinability relations. This is the topic of the present section.

The *relata* are the different putative types of phenomenology. It would be convenient if it were entirely uncontroversial what these are. In reality, an element of theorization is involved even in fixing the inventory of putative phenomenologies. After all, in its concrete unfolding the stream of consciousness

consists in the succession of token phenomenal events and processes, but when we speak of a putative phenomenology, we speak of a phenomenal *type* or *property*. The question arises, then, where these putative phenomenal types or properties come from.

In one respect, the question is not special to phenomenology: in its concrete unfolding, the spatiotemporal universe consists in individual objects and events and the particular ways they are. According to the nominalist about properties, such individuals exhaust what the universe contains. Realists postulate also *universals*, entities simultaneously wholly present in distinct places. Most often, these are construed not as Platonic, transcendent, *ante rem* universals that exist outside space and time and that spatiotemporal entities bear a relation of sharing-in to, but as Aristotelian, immanent, *in re* universals that inhere in the spatiotemporal entities themselves (Armstrong 1978).¹⁸

The epistemology of positing such universals is rather straightforward for Armstrong. When we notice what appear to be objective similarity relations among different concrete particulars, we may infer that there is a universal that they share. This inference is abductive: the fact that completely distinct concrete particulars sometimes resemble calls for metaphysical explanation, and the best explanation is that there is a single entity wholly present in all of them.

The machinery of Armstrongian universals can be applied to the phenomenal realm as well. We may consider that one's stream of consciousness includes in fact not only a succession of token phenomenal states, but also *phenomenal universals* wholly present at temporally disjoint points in that stream (and perhaps in other streams), that is, fully inhering in distinct token phenomenal occurrences. Thus the inventory of putative phenomenologies is just an inventory of putative phenomenal universals—a list of *epistemically possible* phenomenal universals (phenomenal universals there may be).

From this perspective, the epistemology of coming up with the initial inventory of putative phenomenologies is fairly straightforward. It is a matter of seeking phenomenal similarities among individual experiences and inferring from noticed similarities the existence of phenomenal universals shared by them. Presumably, the seeking is through introspection, the inferring through abduction.

Note that this exercise, although straightforward, is doubly fallible. First, one may (perhaps due to attentional overload) mistakenly come to believe that a certain phenomenal similarity holds where in fact it does not, or conversely miss out on a phenomenal similarity that does hold. Secondly, one's abductive inference may be epistemically justified and yet lead to a false conclusion. Still, introspective observing of similarities and abductive inferring from them can be reasonably expected to be *reliable*: although they may lead to false conclusions, they do tend, in the normal go of things, to lead to true ones. (The reliability of introspection will be defended in §§7–8.)

We are now in a position to articulate a *third approximation* of the procedure from §2. This approximation involves decomposing Step 1 into two more basic steps, an introspective step and an inferential step. The former involves introspectively noticing token phenomenal occurrences and recording apparent objective (observer-independent) similarities among them. The latter involves inference from these apparent phenomenal similarities to putative phenomenal universals.



There is a further complication we need to address. It concerns the fact that similarity comes in degrees. There is greater similarity between two experiences as of yellow₁₇ than between an experience as of yellow₁₇ and an experience as of yellow₁₄. The latter similarity is in turn greater than the similarity between an experience as of yellow₁₇ and an experience as of blue₂₃, which in turn is greater than the similarity between an experience as of yellow₁₇ and an auditory experience as of trumpet sounds, which in turn is greater than that between an experience as of yellow₁₇ and a cognitive experience as of $2 + 2 = 4$ (if such there be). One approach to this problem is to accept only *exact similarity* among tokens as indicating the existence of universals. Armstrong himself adopted this view, thus embracing only maximally determinate universals. But this seems to ignore the genuine similarity between inexactly similar tokens (e.g., experiences as of yellow₁₇ and as of yellow₁₄). Another approach is therefore to posit an extra dimension along which universals may differ, depending on the degree of similarity among the particulars they inhere in. Lewis (1983) suggests that universals differ in their degree of *naturalness*: the more objectively similar its instances, the more “natural” the universal. Thus there exist both the universal of phenomenal yellowness and the universal of phenomenal yellow₁₇-ness; the latter is simply more natural than the former.

Happily, the epistemology of establishing the degree of *naturalness* of a phenomenal universal is essentially the same as that of establishing the *existence* of a phenomenal universal. It is a matter of (introspectively) noticing a degree of similarity among token phenomenal states and (abductively) inferring a degree of naturalness in the universal.

Interestingly, this epistemology of naturalness may offer an entry point into the question, raised in the previous section, of how we might establish determinable/determinate relations among putative types of phenomenology. Consider that maximally determinate phenomenal universals are maximally natural: their instances have the greatest possible similarity among them. At the same time, the highest phenomenal determinable, *phenomenality per se*, has the lowest degree of naturalness of all phenomenal universals: its instances are similar only insofar as there is something it is like to have them at all. Thus establishing relative degrees of naturalness in putative phenomenologies may be at least

a partial guide to establishing determinable/determinate relations among them. Using this guide, we can establish that phenomenal yellow₁₇-ness is a determinate of phenomenal yellowness, because (i) the degree of naturalness of the former is higher than the degree of naturalness of the latter (because the former's instances resemble each other more than the latter's), and (ii) instances of the former are also instances of the latter. More generally, for any putative phenomenal universals P and P*, we are justified in taking P to be a determinate of P* if (i) P's instances are similar to each other more than P*'s and (ii) P's instances are also instances of P*.

Setting aside the issue of similarity and naturalness, the third approximation of the overall procedure for producing an account of phenomenal structure would involve five components: 1) introspective noticing of apparent phenomenal similarities; 2) abductive inference from said similarities to putative phenomenal universals; 3) ordering of said universals into layers of phenomenal determinables; 4) identifying the phenomenal primitives in each layer of determinables; 5) identifying the phenomenal derivatives in each layer. The result is a web of phenomenal universals bearing grounding and determinability relations.

5. Realism and Anti-realism about Phenomenal Structure

The picture presented thus far is pleasantly but somewhat dogmatically optimistic in its realism about the various dimensions of structure in the phenomenal realm. It assumes that there are objective, observer-independent facts of the matter about phenomenal similarity, grounding, and determinability relations. Such a pleasantly realist picture may be defensible, but there are also reasonable anti-realist positions on similarity and universals, on grounding, and on determinables.

Start with similarity. Although from our vantage point it is natural to take the similarities we detect around us to be *objective*, it is a recurring idea in the history of philosophy that such similarities may at least sometimes be artifacts of our idiosyncratic cognitive architecture. For example, we can (with difficulty!) envisage a kind of creature C such that instances of green strike C as less inherently similar than instances of grue (Goodman 1954). One might dismiss C's similarity impressions as falsidical, but it would not be easy to fend off a charge of chauvinism here. Although we have a strong sense that finding greater similarity among grue than green instances is perverse, this may simply be because we are insufficiently imaginative. For the anti-realist about similarity, there is no fact of the matter as to which group of instances is *really* (objectively, absolutely, inherently) more similar (Gärdenfors 2004). According to the realist, there are—they are the naturalness facts mentioned in §4. This debate is as old as philosophy itself, and this is not the place to make meaningful contributions to it. I mention it merely to bring up the epistemic possibility that the phenomenal

similarity relations underlying our carving of the phenomenal realm into putative phenomenal universals may be observer-dependent.

One could reach such anti-realism about phenomenal similarity by simple application of global anti-realism. Alternatively, one might consider there to be something special about the phenomenal realm that lends it to anti-realist treatment. For example, one might hold that, in the phenomenal realm more than elsewhere, some similarity disagreements may be “blameless.” Imagine three phenomenal tokens φ_1 , φ_2 , and φ_3 , of which two counterfactuals are true: (a) if they occurred in Luciana’s stream of consciousness, it would introspectively seem to her that φ_1 and φ_2 resemble each other but do not resemble φ_3 ; (b) if they occurred in Ada’s stream of consciousness, it would introspectively seem to her that φ_2 and φ_3 resemble each other but not φ_1 . Let us further stipulate, as we seem entitled, that Luciana and Ada are both healthy adults with well-functioning introspective faculties, and that these similarity judgments were issued in favorable epistemic circumstances. It is at least antecedently reasonable to maintain that, at least in some cases fitting this description, neither Luciana nor Ada is wrong: φ_2 resembles φ_1 relative to Luciana’s introspective apparatus but φ_3 relative to Ada’s. This would suggest that phenomenal similarity relations are introspection-dependent.¹⁹

The resulting anti-realism could play out in different ways. One view might be that phenomenal tokens φ_j and φ_k resemble iff an ideal introspector would (under ideal conditions) introspectively take them to. Another might be that φ_j and φ_k resemble iff normal introspectors (under normal conditions) do introspectively take them to. Other variants are possible. What they would all have in common is a certain *a priori* tie between phenomenal similarity and introspective impression thereof.



It is not my purpose here to settle any debate over realism and anti-realism about phenomenal similarity. But note that the debate has implications for the status of phenomenal universals: whether they are observer-dependent or -independent. A full theory of the structure of the phenomenal realm should take a position on the matter.²⁰ It would also take a position on a similar debate about determinability relations. The discussion above assumed that such relations are fully objective. But a more conventionalist or pragmatist approach would not be unreasonable here. On such an approach, the classification of phenomenal universals into determinacy layers is not exactly a matter of tracking observer-independent facts; rather, it is a subtler exercise involving choice of a *useful* scheme to adopt (where the usefulness of a scheme is relative to interests, goals, and so on). In a particularly radical form, the idea might be that a “correct taxonomy” of phenomenal properties is not waiting to be *discovered*, it is waiting to be *created*. Hybrid options would take classification to involve an elusive

admixture of discovery and creation. These hybrids cast classification as *partly* observer-dependent.

Consider an analogy. It is apparently a matter of some contention whether Anglican Christianity should be considered a species of Protestantism, a species of Catholicism, or a *sui generis* branch. The first two views classify Anglicanism as an n -layer determinable, but the third as an $n-1$ -layer determinable. In this case, it is highly plausible that the choice between the three classificatory schemes is not *just* a matter of accurately tracking some observer-independent facts (historical, doctrinal, or other). There are almost certainly pragmatic and conventional considerations that bear on the choice. It is perfectly respectable to hold that some issues pertaining to phenomenological taxonomy follow the same model. At the same time, a more realist or objectivist approach would be perfectly respectable as well. According to such realism, the phenomenal realm has natural joints and the purpose of phenomenological taxonomy is to capture those accurately. It is by no means immediately obvious which approach is more plausible.

This antecedent uncertainty trickles down to the realm of maximally determinate phenomenal properties—the “phenomenal elements”—such as phenomenal-yellow₁₇-ness. Some of the latter’s instances occur on the weekend, some during the week. Some occur at night, some during daytime. So the putative properties of weekday-phenomenal-yellow₁₇-ness, nightly-phenomenal-yellow₁₇-ness, and so on are *more* determinate than just phenomenal-yellow₁₇-ness. Phenomenal-yellow₁₇-ness is a genus of which they are species. Moreover, the similarity among all instances of weekday-phenomenal-yellow₁₇-ness cannot be lower than that among all instances of phenomenal-yellow₁₇-ness. Nonetheless, it seems absurd to deny phenomenal-yellow₁₇-ness its status as a phenomenal element for this reason. The realist and the anti-realist offer different explanations of this absurdity. The realist claims that phenomenal-yellow₁₇-ness is more *natural* than weekday-phenomenal-yellow₁₇-ness. The anti-realist maintains that it is much more *useful* to us, given our interests and purposes.²¹ The realist explanation is perhaps more initially attractive, but faces difficulties. In particular, if the realist claims that phenomenal-yellow₁₇-ness is more natural than weekday-phenomenal-yellow₁₇-ness, she must offer some criterion of naturalness that goes beyond similarity. Thus an anti-realist take on what makes a certain phenomenal property “elemental” is far from incredible.



Some grounding relations, such as mereological composition, have sometimes been claimed to be contingent (Rosen 2006, Cameron 2007). Suppose for the sake of argument that the grounding relations among phenomenal properties are likewise contingent. Then there should be distinct metaphysically possible

worlds which are qualitatively indistinguishable but which differ with respect to what grounds what. Take any (same-layer) phenomenal universals P_1 , P_2 , and P_3 . If phenomenal grounding is contingent, then there are three otherwise similar possible worlds W_1 , W_2 , and W_3 , such that: in W_1 , P_1 and P_2 are primitive, while P_3 is grounded in their combination; in W_2 , P_2 and P_3 are primitive, while P_1 is grounded in their combination; in W_3 , P_1 and P_3 are primitive, while P_2 is grounded in their combination. But this seems a little odd. There is some intuition, I think, that in fact W_1 , W_2 , and W_3 are one and the same world, differently described. Since W_1 , W_2 , and W_3 contain all the same particulars, and moreover these particulars have all the same nonrelational properties, they are in truth indistinct worlds. This may well be the view of some meta-metaphysical anti-realists (see Sidelle 2002). However, there are also capable defenses of meta-metaphysical realism that would insist on an objectivist take on grounding and would treat W_1 , W_2 , and W_3 as genuinely different worlds (Schaffer 2009, Sider 2011).²² As before, I raise the issue only to register the multiplicity of possible approaches to the project of understanding phenomenal structure.²³

In conclusion, phenomenal grounding, determinacy, and similarity each admit of both respectable realist and respectable anti-realist positions. It is even coherent to combine realist and anti-realist positions on different aspects of phenomenal structure (for example, realism about phenomenal similarity with anti-realism about phenomenal grounding and determinability). The most desirable view in this area is robust realism about phenomenal similarity, determinability, and grounding alike. It paints forth an upliftingly objectivist picture of the phenomenal realm as fully and intricately structured in and of itself, suffused with observer-independent structure awaiting discovery. However, it is not improbable that some aspects of phenomenal structure (understood as comprising, *inter alia*, similarity, determinability, and grounding relations) might turn out to partly express our introspective, cognitive, pragmatic, and other predilections. Ultimately, a full theory of the structure of the phenomenal realm would have to take a position on this matter. In this book, though, I stay neutral on this issue and pursue the aspects of phenomenal structure pertinent to the question of primitiveness independently of whether they are observer-(in)dependent.

II. Methodological Foundations

6. The Role of Introspection

If you try to understand the structure of your car's carburetor, you start by *looking* at it. It is extremely hard to make much progress on understanding the interrelations among the carburetor's various parts without looking—a blind person

would find the task extremely challenging. Thus *perceptual encounter* with the structure is the starting point for any plausible attempt to decipher it.

Fortunately, the carburetor's concrete physical structure lends itself to perceptual encounter. When a structure is abstract, or mental, it does not lend itself to perceptual encounter. We then hope for some other type of encounter with it. In describing his experience of attempting to understand mathematical structures, the Polish logician Jan Łukasiewicz (1970: 249) extols the virtues of *intuitive encounter*:²⁴

I should like to sketch a picture connected with the deepest intuitive feelings I always get about logistic . . . Whenever I am occupied even with the tiniest logistical problem, e.g. trying to find the shortest axiom of the implicational calculus, I have the impression that I am confronted with a mighty construction, of indescribable complexity and immeasurable rigidity. This construction has the effect upon me of a concrete tangible object, fashioned from the hardest of materials, a hundred times stronger than concrete and steel. I cannot change anything in it; by intense labor I merely find in it ever new details, and attain unshakable and eternal truths.

If concrete physical structures lend themselves to perceptual encounter and abstract structures offer a sense of intuitive encounter, concrete phenomenal structures may present themselves to *introspective encounter*. Thus in attempting to understand the structure of one's stream of consciousness, it is natural to start by introspectively observing it. On the face of it, understanding the stream's phenomenal structure without introspecting it would be as challenging as understanding a carburetor's physical structure without looking at it.

Unlike perceptual encounter, the notions of intuitive and introspective encounter have faced sustained resistance in much twentieth-century philosophy. Let us set aside intuition, which does not bear directly on our present concerns. When it comes to introspective observation, the very notion that introspection should be understood along observational lines, on the model of perceptual observation, has often been challenged. As I have offered a sustained argument for the observational model elsewhere (Kriegel 2011b ch.1), here I merely wish to draw implications for the study of phenomenal structure. In particular, I want to argue that the model casts introspection as indispensable for such a study.

Consider other areas of inquiry. Zoologists who study zebras pay close attention to the impact of zebras on their environments, the environment's impact on zebra populations, various correlates and indicators of zebra presence, and so on. But in constructing their theories of zebras, zoologists use not only observations of such zebra-indicators. They also use observations of *zebras*. Indeed,

given that it is possible to observe zebras, it would be folly for zoologists to refuse to take into account observations of zebras in constructing their theories of zebras. When studying leptons, we must construct our theories without taking into account direct observations of leptons, since we *cannot* observe the leptons themselves. But given that we *can* observe zebras, it would be perverse to construct our theories thereof without taking zebra observations into account. More generally, whenever we *can* observe a type of phenomenon, it is perverse to insist on developing our understanding of it in complete disregard of our observations of tokens.

This general principle has immediate implications for our understanding of phenomenal consciousness. If some conscious phenomena *can* be observed, it would be very odd indeed to insist on bracketing all such observation in developing our mature theories of consciousness. This suggests a straightforward argument for the *epistemic indispensability* of introspection. Let us say that an understanding of P is legitimate just when it is constructed or arrived at in an epistemically responsible manner. Then:

- 1) When a phenomenon P is observable, any legitimate understanding of P must take account of observations of P;
- 2) Some conscious phenomena are introspectively observable (and not otherwise observable); therefore,
- 3) Any legitimate understanding of consciousness must take account of introspective observations of conscious phenomena.

In the remainder of this section, I want to defend this argument from two sources of resistance. From this will emerge a fuller defense of the role of introspection in the contexts of discovering and justifying phenomenological hypotheses.

Before starting, it is worth pausing to draw a distinction between narrow and wide conceptions of introspection. The most straightforward mode of first-person awareness of one's own experience involves turning one's attention inward and attending to one's own concurrent internal goings-on. We may call this "introspection proper." But distinctly first-person awareness may be wider in two ways. First, it may be possible to conjure up a past experience in episodic memory and a future or merely possible experience in imagination, and then use broadly introspective capacities to examine these remembered or imagined experiences. Titchener (1912) called this "indirect introspection." Secondly, it has sometimes been claimed that some or all of our experiences are accompanied by a kind of nonattentive, unimposing, "peripheral" inner awareness (Brentano 1874, Kriegel 2009). To the extent that introspection proper is always attentive, this other inner awareness is not properly introspective. But it is a distinctly first-personal awareness nonetheless: nobody else can have *this* kind of

awareness of my own experiences. It may thus be useful to collect under a single heading (i) this nonattentive inner awareness, (ii) introspectively aided examination of remembered and imagined experiences, and (iii) introspection proper. We may use the label “introspection loosely so called” for this wider group of capacities. Most of what I say below is intended to apply to introspection loosely so called. For this reason, when I do not explicitly specify which notion of introspection I am using, I should be understood to speak of introspection loosely so called.

6.1. *Introspective Discovery*

A straightforward objection to introspective indispensability is that the history of cognitive science gives the lie to it. Introspection cannot be epistemically indispensable, since cognitive science has actually dispensed with it.

To my mind, there are two problems with the idea that cognitive science is doing fine without introspection. The first is that it is not doing fine. The second is that it is not doing without introspection. It is a familiar comment—though admittedly a controversial one—that cognitive science has met with limited success when it comes to phenomenal consciousness.²⁵ This is the sense in which cognitive science is not doing fine without introspection, and I will not belabor the point here. Let me focus instead on my claim that cognitive science is not doing without introspection to begin with.

To appreciate the enduring role of introspection in cognitive science, let us start with the distinction, often attributed to Reichenbach but present already in Bolzano (1837 §15), between the “context of discovery” and the “context of justification.” A dramatic example is provided by the German chemist Friedrich Kekulé’s discovery of the molecular structure of benzene. The evidence Kekulé (1865) cited in justification of his model of benzene had to do with isomers, derivatives, and so on. But a quarter-century later Kekulé recounted that his initial discovery was due to a daydream in which he “saw” a snake biting its own tail. Clearly, in this case, the manner in which the model was discovered is irrelevant to the manner in which it is justified. Thus the contexts of discovery and justification can come apart, even if typically the gap between them is not this dramatic.

My claim in this subsection is that even *if* introspection has been purged from the context of cognitive-scientific justification, it certainly continues to underlie large tracts of research in the context of cognitive-scientific discovery. In fact, significant portions of modern cognitive science strike me as based on introspective discovery paving the way to non-introspective justification. Often the scientist, being a reflective introspector, experiences an initial introspective insight into some psychological phenomenon, and on its basis forms a hypothesis; s/he then proceeds to devise experimental tasks that ingeniously use exclusively

third-person measures (often reaction times) to generate non-introspective evidence for the introspectively formed hypothesis.

A fine example of this is Roger Shepard's seminal work on imagery and the phenomenon of mental rotation (Shepard and Metzler 1971). That we use mental rotation of private images to compare shapes of the objects imaged is of course what introspection teaches—there is nothing surprising there. The ingenuity in Shepard's research was in devising an experimental paradigm in which pairs of similarly shaped but differently oriented three-dimensional "objects" with varying angles of putative rotation were to be judged for similarity by subjects. The fact that subjects took longer to judge the shapes to be similar when the angle of putative rotation was greater suggested that these subjects were mentally rotating imagistic representations of the objects. Thus the purely "objective" (read: third-person) measure of reaction time served to ratify what was already known on the basis of "subjective" (first-person) introspective impression.

A more recent example is Vilayanur Ramachandran's ingenious demonstration of number-color synesthesia (Ramachandran and Hubbard 2001). Ramachandran used panels of numerals printed in a way that made it difficult to distinguish different numerals. Control subjects took significantly longer to identify incongruent numerals than number-color synesthetes, to whom the incongruent numerals presumably appeared incongruously colored. Of course, that number-color synesthesia exists we know on the basis of synesthetes' introspective reports since at least the nineteenth century (see Galton 1880). But Ramachandran's reaction-time-based demonstration had the advantage of purging appeal to introspection in ratifying this knowledge.²⁶

From casual observation of cognitive-science conferences and colloquia, my impression is that this sort of gambit is pervasive in vision science and throughout cognitive (neuro-)psychology. Scientists often devise ingenious experimental designs that circumvent explicit appeal to introspection, but the original hunch underlying the research is founded on personal introspection.²⁷

It is an open question just how far cognitive science would get if it purged introspection not only from the context of justification but also from the context of discovery. Suppose cognitive science insisted, from its inception, not only on devising non-introspective justification of introspectively formed hypotheses, but also on exclusively non-introspective hypothesis formation. In fact, consider a possible world otherwise like ours but where cognitive scientists lack any introspective capacities. My own suspicion is that we would be shocked to find out just how skeletally poor the scientific understanding of the mind is in such a world. If so, the role of introspection in our own scientific understanding of the mind is greatly underrated in the "official narrative" about cognitive science. For the gap between the state of our knowledge and understanding and the state of knowledge and understanding in that counterfactual world is owed entirely to our implicit ("unofficial") use of introspection.

If all this is right, then introspection is not only epistemically indispensable, but also not really dispensed with in cognitive-scientific practice. Still, introspection skeptics may insist that introspection has no role to play in the context of cognitive-scientific justification. This is the topic of the next subsection.

6.2. Introspective Justification

To be sure, there is a long philosophical tradition of over-trusting introspection. In its strongest form, this tendency can be articulated as a conjunction of two converse theses, one asserting the *perfect reliability* of introspection and one the *omnipotence* of introspection. According to the first, introspection is *infallible*:

(II) If subject S introspects having phenomenology P, then S has P.²⁸

According to the second, introspection never misses anything that passes within its purview, rendering phenomenology *self-intimating*:

(SI) If subject S has phenomenology P, then S introspects having P.

The conjunction of II and SI casts introspection as perfectly trustworthy. We may call the conjunction *introspective dogmatism* (or perhaps *introspective maximalism*, since it portrays introspection as maximally powerful).

Unfortunately for all involved, introspective dogmatism (or maximalism) is highly implausible. Introspection is far from perfectly reliable and far from omnipotent (Nisbett and Wilson 1977). However, the fact that introspection is not maximally trustworthy does not show that it is thoroughly unreliable and/or entirely impotent.²⁹ For our present purposes, what matters is whether introspection could be shown to be *minimally* trustworthy, that is, have the least demanding epistemic properties that would be needed for it to play a legitimate role in the context of justification. This requires that we identify these minimal epistemic properties, concerning both reliability and potency.

On the side of reliability, plausibly what is required is that introspection enjoy *above-chance reliability*: it is more likely that one has a phenomenology if one introspects having it. On the side of potency, being minimally justificatory would plausibly require that introspection enjoy *nonnegligible potency*: it is more likely one will introspect a phenomenology if one has it. To a first approximation, we may formulate the claim as follows:

(ACR) If subject S introspects having phenomenology P, then S is more likely to have P than if S does not so introspect.

(NNP) If subject S has phenomenology P, then S is more likely to introspect having P than if S does not have P.

The mere *disjunction* of ACR and NNP already bestows *some* epistemic value on introspection. A more robust yet highly plausible view consists in their conjunction; call it *introspective minimalism*. The idea is that having a phenomenology makes it more likely that one introspects it and introspecting it makes it more likely that one actually has it. I contend that this view, or something very like it, undergirds the legitimacy of appeal to introspection in *justifying* hypotheses about consciousness. Something stronger may yet be true, but the truth of minimalism would suffice to legitimize introspective appeal.

Suppose, for instance, that introspection turned out to be as trustworthy as our sense of smell, that is, as reliable and as potent as a normal adult human's olfactory system. Then introspective minimalism would be vindicated. Normally, when we have an olfactory experience as of raspberries, it is more likely that there are raspberries in the vicinity than if we do not have such an experience. Conversely, when there are raspberries in the vicinity, it is more likely that we would have an olfactory experience as of raspberries than if there are none. So the "equireliability" of olfaction and introspection would support introspective minimalism. Such equireliability is highly plausible.

It is worth noting that introspective minimalism can be refined in various ways. Thus, ACR and NNP do not explicitly contain any quantifiers, suggesting that they are intended as doubly universal, applying to *all* subjects and *all* phenomenologies (all values of S and P). This may turn out to be too strong. Perhaps it would be wiser to restrict these claims to normal subjects normally circumstanced. For it may be that under conditions of cognitive overload, or in psychologically malformed subjects, introspecting a phenomenology does not increase the probability that the phenomenology is present (and/or the presence of a phenomenology does not increase the probability that it be introspected).³⁰ Likewise, there may be reasons to exclude certain special types of phenomenology from ACR or NNP. Various "exemptions" may need to be carved. For example, according to many phenomenologists, a person's field of consciousness typically involves a "fringe" or "margin" that contributes to one's overall experience very lightly and unimposingly. Thus, as I sit in front of my laptop and visually experience it in an attentive and focused manner, I am also aware, much more peripherally and as it were almost imperceptibly, of the tactile sensation of soles of my shoes, a low-humming anxiety about a looming appointment with a plagiarizer, and so on. Arguably, this fringe phenomenology cannot in principle be introspected, since introspecting a phenomenal occurrence renders it focal rather than fringe (Kriegel 2009 ch.5). Likewise, consider the phenomenology of experiential immersion or engrossment, such as a basketball player experiences when "in the zone." This immersive phenomenology may also be non-introspectible, insofar as turning one's introspective attention onto it would require taking a step back from it and disrupting its characteristic feelings of rightness and flow.³¹ More generally, there may be a class of phenomenologies

whose very essence requires the absence of introspective attention; we may call these (doubtless suboptimally) “elusive phenomenologies.” If so, ACR and NNP would probably need to be restricted to nonelusive phenomenologies.³²

Taking into account the just-discussed restrictions, we would obtain the following doubly refined thesis of introspective minimalism:

(RIM) For any (normally circumstanced) *normal* subject S and any *nonelusive* phenomenology P: If S introspects having P, then S is more likely to have P (than if S does not so introspect) & If S has P, then S is more likely to introspect having P (than if S does not have P).

Other restrictions may be called for upon closer examination. Still, the fully refined minimalist thesis would very likely be nontrivial yet in a position to undergird the legitimate scientific use of introspection not only in the context of discovery but also in the context of justification.

7. The Problem of Introspective Disagreement

It is sometimes felt that the most important challenge to introspective appeal is not so much the epistemic status of an individual’s introspective impressions, but rather the apparent irresolvability of interpersonal introspective disagreements. Suppose S reports her introspective impression that phenomenology P has feature F, whereas S* reports that P does not have F. (We may further suppose that S and S* report the same degree of confidence in their introspective impressions and both have properly functioning introspective faculties.) How is our investigation of P supposed to proceed under such (not uncommon) circumstances?

The question is pressing, because arguably it was precisely the problem of introspective disagreement that has led to the demise of introspectionist psychology. Two disagreements between Titchener’s lab and Külpe’s have proved particularly intractable, and in the long run particularly damaging to introspectionism. One concerned the existence of imageless thoughts: for Titchener and his students, competent introspection revealed that there are none; for Külpe and his, it revealed there are some. The other concerned the number of phenomenal elements: Titchener required 42,415; Külpe managed with only 11,000 (Revonsuo 2010 ch.2). It is this type of apparently irresolvable dispute that has led Watson (1913: 163) to dismiss introspectionist psychology wholesale:

Psychology, as it is generally thought of [i.e., by introspectionists], has something esoteric in its methods. If you fail to reproduce my findings, it is not due to some fault in your apparatus or in the control of your stimulus, but it is due to the fact that your introspection

is untrained . . . In [the natural] sciences a better technique will give reproducible results. Psychology is otherwise. If you can't observe 3–9 states of clearness in attention, your introspection is poor. If, on the other hand, a feeling seems reasonably clear to you, your introspection is again faulty. You are seeing too much. Feelings are never clear.

Today, similar disagreements bedevil debates over cognitive phenomenology, emotional phenomenology, and so on. One is entitled to worry, therefore, that current phenomenological disputes will prove no more tractable, and in the final analysis no more fruitful, than the original introspectionists'.

In this section, I want to consider three possible reactions to such a predicament. I will call them the *no-fact* view, the *phenomenal variability* view, and the *introspective competence* view. These reactions can be adopted in different circumstances and can be more appropriate in some cases than others. I will argue that the last type of reaction is the most appropriate for the kind of cases at issue in this book.

The most staunchly defeatist reaction to introspective disagreements is that there is *no fact of the matter* as to what the phenomenology is *really* like. In the present case, there is no determinate phenomenological fact that settles the question of whether P is F or not. (This is not to be confused with the eliminativist claim that there is no such thing as P, or F. Such eliminativism does provide for facts that settle the relevant questions.) Call this the *no-fact view*.³³

The no-fact view can play out in two different ways, depending on whether we adopt descriptivist or expressivist semantics for introspective reports. The more natural version of the view is descriptivist, and claims that both S and S* are attempting to correctly describe the phenomenological facts when they report "P is F" and "P is not F." The upshot is an introspective error theory: since there are no such phenomenological facts, both reports are untrue.³⁴ There is, however, an expressivist version of the view, according to which the disagreement between S and S* is *merely apparent*, since neither is actually making an assertion. Instead, each expresses a different (noncognitive) attitude toward P's being F: S approves of P being F, S* disapproves.³⁵ In both versions of the no-fact view, introspective reports do not constitute a source of evidence for theorizing about consciousness.

The main problem with the no-fact view is that, pending further considerations, it comes across as an *overreaction* in virtually every case. In other areas, we do not conclude from the existence of disagreement that there must be no fact of the matter. If S contends that she saw the keys in the living room but S* insists that she saw them in the dining room, we would be disinclined to conclude that there is no fact of the matter as to where the keys really are. So it cannot be just the *existence* of disagreement that leads to the no-fact view. Perhaps what is special about introspective disagreement is that they persist through the

end of inquiry, whereas disagreements on key locations tend not to. The problem with this is that there is no meaningful evidence for the claim that introspective disagreements will persist through the end of inquiry. One might suggest that they would *have* to, given the incorrigibility of introspection.³⁶ But, setting aside the fact that the incorrigibility thesis appears too strong to be plausible, even conceding it could at most lead to an *epistemological* conclusion, namely, that we cannot *know* whether P is F or not. Nothing here entails the stronger *metaphysical* conclusion that there is no fact of the matter as to whether P is F.³⁷



A more cheerful reaction to introspective disagreements is what we may call the *phenomenal variability view*. According to it, even the most fundamental features of phenomenology often vary across subjects. The right attitude to our disagreement is to conclude that S's P is F but S*'s it is not F. Their phenomenologies are simply different.³⁸

This view may take inspiration from perceptual disagreements. Suppose S and S* are each facing a swan, and S reports that the swan before her is white while S* reports that the swan before her is black. If we know that the swans perceived are numerically different, it would be most natural to conclude that the swans are also qualitatively different: S's is white but S*'s is black. In the phenomenal case, the experiences introspected are definitely numerically different (since one occurs in S and one in S*). So it should be likewise natural to conclude that S's P-experience is F while S*'s is not. This reasoning only requires the plausible premise that introspection is highly reliable. If we know (i) that in S and S* introspection is highly reliable, (ii) that S introspects one instance of P and S* another instance, and (iii) that S's introspective impression is that P is F and S*'s is that P is not F, then we should infer that S's instance of P is F but S*'s is not F.

The phenomenal variability reaction is surely the right one in many cases of introspective disagreement. However, the more basic and general the phenomenal feature at issue, the less plausible the view becomes. If S claims that her phenomenology of frustration involves a tinge of anger, whereas S* claims that hers does not, it is fitting (*mutatis mutandis*) to adopt the variability reaction. But suppose S reports that her emotional phenomenology involves nothing more than feeling certain bodily sensations, whereas S* reports that hers involves an additional element that constitutes a *sui generis* emotional phenomenology. It would be odd to conclude that S and S* simply differ in their emotional phenomenology, one's being primitive while the other's reducing to bodily sensations. That is, it would be odd to conclude that the James-Lange theory of emotion is true of S but false of S*.³⁹ We would be more inclined to think that S and S* must have conceptualized differently the same phenomenology, or that one of them did not fully understand what was at issue, or that one committed an introspective mistake or oversight. Likewise if S claimed to have phenomenal

properties at all while (otherwise normal) S* professed to be a zombie: a let-a-thousand-flowers-bloom approach would seem misplaced here.

My focus in this book is on second-layer phenomenal determinables. These are the most general phenomenal features other than phenomenality *per se*. To that extent, I am interested in phenomenal features for which the phenomenal variability approach is least appealing. For this reason, I will set it aside here as unlikely to affect the issues to be discussed.



According to the no-fact view, both S and S*'s introspective reports are untrue. According to the phenomenal variability view, both are true. A third option is that one report is true and the other untrue. On this view, which we may call the *introspective competence* view, introspective disagreements show nothing more than that one of the parties to the disagreement must be wrong. Thus, if S and S* make incompatible introspective judgments on the same type of phenomenology, the right response is to simply seek who is in error.

With perceptual disagreements, this is often the most natural approach. Suppose S and S* smell a bottled extract, and S judges it is gardenia while S* judges it is jasmine. Our natural inclination would be to suppose that one of them must be simply mistaken. One of them is the better nose, the more competent odor-detector. Likewise, at least when subjects introspect the most general features of phenomenology—those for which phenomenal variability is least plausible—it should be natural to suppose that disagreements are due to differences in introspective competence. When S and S* issue conflicting introspective reports (on a very general aspect of phenomenology), the theorist would be justified, *mutatis mutandis*, in endorsing the more competent introspector's report. Such endorsement may result in error, of course, in case the more competent introspector got it wrong. But on the whole, higher competence should be the more reliable guide. So, endorsing competent introspectors' reports would always be epistemically justified, at least by reliabilist lights.

We may think of this as a heuristic for investigating phenomenal structure. *Applying* this heuristic does require that we have a competence measure, one independent of which introspector we theorists happen to agree with. Admittedly, such a measure should be hard to devise. But once it is, introspective disagreements submit to straightforward resolution: we are to endorse whichever introspective report is issued by the more competent introspector.

There is certainly a firm pretheoretic resistance to this approach. But the resistance may be more moral than epistemological. Charles Siewert once pointed out to me that our privileged access to our own phenomenology is a (“the”?) central source of our sense of dignity as separate, inviolable, self-possessing individuals.⁴⁰ This is why telling people what they *really* feel, overriding their own claims about what they feel, seems first and foremost *morally abrasive* rather

than *epistemically irresponsible*.⁴¹ Conversely, being told what one really feels over one's protestations tends to elicit moral rather than epistemological indignation. The full case for this would have to be prosecuted elsewhere, but my suspicion is that the ethos of first-person incorrigibility is rooted in this sort of respect for the inviolable dignity of others.⁴² It is this ethos, I am suggesting, that accounts for our pretheoretic resistance to the introspective competence view. The view sanctions a domineering attitude toward overridden subjects—which makes us cringe. Conversely, there is a feel-good element to the phenomenal variability view: it respects persons. Nonetheless, we must keep in mind that these pressures away from the introspective competence view are not properly epistemic. They do not provide us with *epistemic* reasons for preferring the phenomenal variability view, only *moral* reasons.

This line of thought effectively provides a *debunking explanation* of our pretheoretic resistance to the introspective competence view (and attraction to the phenomenal variability view).⁴³ All the same, the competence approach is more helpful in principle than in practice. For in practice we have no idea how to devise a sufficiently compelling competence measure. Ideally, the measure would be so compelling that it would be possible for a theorist, who recognized that (a) her introspective impressions conflicted with a peer's and (b) the peer scores higher on the relevant measure, to let the peer's reported impressions override her own. It is at present unimaginable what such a measure might look like. Wundt reportedly considered an introspector competent only after 10,000 trials (Schwitzgebel 2011: 74). But this does not take into account how attentive and alert the subject is in each trial, how antecedently talented (perceptive and reflective) an introspector she is, and myriad other factors. Yet no better competence measure suggests itself.⁴⁴ In addition, we have no clear standards for when to adopt the introspective competence approach to begin with. It was noted above that the approach is more plausible than phenomenal variability when the phenomenal feature at issue is sufficiently general, that is, constitutes a sufficiently high-level determinable. But this does not yet give us a concrete threshold of generality/determinacy to guide choice between phenomenal variability and introspective competence in practice.

To conclude, I have considered three possible reactions to introspective disagreements. The no-fact view casts the introspective reports of both parties to the disagreement as untrue, the phenomenal variability casts both as true, and the introspective competence view casts one as true and the other as untrue. The correct approach seems to me to be this: when the disagreement is over a relatively specific phenomenal feature, the phenomenal variability view is *prima facie* more plausible; when the disagreement is over a relatively general phenomenal feature, the introspective competence view becomes more plausible (at least *prima facie*).⁴⁵ Unfortunately, however, this provides only the vaguest guide for dealing with introspective disagreements in practice. Methodological

developments in this area may yet be possible that would blunt the skeptical effect of introspective disagreements. But until such time we must seek ways to limit or eliminate the potential for introspective disagreement. This is the topic of the next section.

8. Beyond Introspection

Introspective disagreements are especially paralyzing if introspection is considered the end-all and be-all of disputes about the phenomenal facts. But often appeal to introspection is only a first step in theorizing. Reasoning, inference, and argumentation are intended to carry the bulk of the case for or against some phenomenological thesis. The phenomenological thesis is not simply *delivered* by introspection: its proponent does not profess to have directly introspected its truthmaker. Rather, the phenomenological thesis is *suggested* by introspection, in the sense that what introspection delivers can be used in an *argument* for it.

Specially prominent in recent phenomenological discussions have been so-called phenomenal contrast arguments. A classic example is an argument for cognitive phenomenology due originally to Moore (1953) and developed later by Strawson (1994 ch.1). Some philosophers avow direct introspective awareness of a purely nonsensory phenomenology special to cognition; others disagree. The Strawson-Moore argument attempts to provide *indirect* support for the existence of such phenomenology. Imagine two subjects S and S*, such that (i) S and S* are listening to the news in French, (ii) S understands French, and (iii) S* does not understand French. In this scenario, it seems that there is a difference in the *overall* phenomenology of S and S* as they listen to the newscast, but that their strictly sensory phenomenology is the same. It would follow that the difference in their overall phenomenology must be due to some nonsensory phenomenology, presumably a purely cognitive phenomenology of *understanding* present in S's experience but not in S*'s. The argument may be reconstructed as follows:

- P1) S and S* have a different overall phenomenology;
- P2) S and S* have the same sensory phenomenology; therefore,
- C1) There is a purely nonsensory phenomenology that S and S* differ in;
and therefore,
- C2) There is such a thing as purely nonsensory phenomenology.⁴⁶

The ultimate conclusion of this argument (C2) is a phenomenological claim on which there is clearly no introspective agreement. But the hope is that the phenomenological premises (P1 and P2) are introspectively uncontroversial *in comparison*.⁴⁷ Thus an argument can provide indirect introspective support for a phenomenological proposition for which direct introspection has failed to produce widespread agreement.

Notice that the Moore-Strawson argument does appeal to introspection in supporting an irreducible cognitive phenomenology. However, it does not appeal to introspection *of* an irreducible cognitive phenomenology. Instead, it appeals to what it hopes are introspectively less controversial claims. The general strategy, then, is this. When a phenomenological proposition q is met with (wide) introspective disagreement, we seek another phenomenological proposition p , such that (i) p does not tend to induce (as much) introspective disagreement and (ii) there is some argument A whose conclusion is q and whose only phenomenological premise is p .⁴⁸ What introspection delivers in A is only p , the relatively uncontroversial phenomenological claim. The route from p to q is not delivered by introspection, but by reasoning. In this fashion, the proponent and opponent of q can isolate some introspective common ground p and debate only the plausibility of $q|p$, that is, q -given- p (with introspection playing no role in that part of the debate). This represents, in effect, a fourth possible reaction to introspective disagreements: the theorist remains neutral with respect to introspection's verdict on the phenomenological proposition at stake, and instead seeks a deductive argument with which to bypass direct appeal to introspection. We may call this the *deductive-bypass approach*. (Although the argument leading from p to q need not be deductive, it *may*.)

The obvious drawback in the deductive-bypass approach is that it still makes *some* introspective appeal. Although the proponent of a deductive argument for a phenomenological thesis does not profess to directly introspect the truthmaker of the argument's conclusion, she does profess to directly introspect the truthmaker(s) of at least some premise(s). This is unproblematic as long as the relevant premises succeed in commanding wide introspective assent. But whether they do is not in the proponent's control. The Moore-Strawson argument, for example, has met with significant resistance to P1 and especially P2. Many philosophers have claimed that they do *not* introspect the sameness of sensory phenomenology asserted by P2 (Carruthers 2006, Robinson 2006, Prinz 2011). This does not mean, of course, that such argumentation is "epidialectical": as long as the phenomenological premises garner *more* introspective assent than the phenomenological conclusion, the argument as a whole represents dialectical progress (other things being equal).⁴⁹ It produces new dialectical pressure on the opponent.⁵⁰ Still, there is a sense in which the possibility-in-principle of a dialectical impasse continues to loom.



An ideal solution to this problem would be to devise arguments for phenomenological conclusions that do not involve introspection at any stage. Unfortunately, it is highly implausible that any collection of non-phenomenological propositions could entail a phenomenological proposition. (A proposition is phenomenological just when its truth requires that some phenomenal property be

instantiated.) If so, every deductive argument for a phenomenological conclusion would have to involve some phenomenological premise(s). This can be thought of on analogy with Moore's (1903) principle that every deductive argument with a normative conclusion must have a normative premise, i.e., that you cannot derive an "ought" from an "is." To salute Moore's notion of "naturalistic fallacy," I will call its phenomenological parallel the *physicalistic fallacy*:

(PF) A deductive argument with a phenomenological conclusion must have at least one phenomenological premise.

Note well: PF in no way requires us to reject physicalism as such. It only requires us to deny that there are *a priori* connections between physical propositions and phenomenal propositions. That is, it requires us to reject "*a priori* physicalism," what Chalmers (2002) calls "type-A physicalism." PF is perfectly compatible with *a posteriori* or "type-B" physicalism, according to which there are necessary but merely *a posteriori* connections between physical and phenomenal facts.⁵¹ In what follows, I will assume that *a priori* physicalism is implausible, and concomitantly that PF is plausible.⁵² A consequence is that no deductive argument for a phenomenological thesis could proceed without phenomenological premises, and therefore (ultimately) without appeal to introspection.⁵³

Importantly, PF does not rule out the possibility of introspection-free arguments for phenomenological theses. It rules out only *deductive* arguments of the sort. It leaves the door open to nondeductive, non-demonstrative arguments. In the recent literature, some *epistemic arguments* have been offered that employ no phenomenological premises but argue through inference to the best explanation for phenomenological conclusions. Perhaps the best known, presented first by Goldman (1993) but developed more fully by Pitt (2004), targets irreducible cognitive phenomenology as well. It proceeds by claiming that our knowledge of some of our cognitive states has certain characteristics that can only be explained, or at least are *best* explained, by the hypothesis that these cognitive states exhibit an irreducible type of phenomenology. Pitt's version of the argument may be reconstructed as follows:

- P1) Subject S has immediate, noninferential knowledge of some cognitive states S is in;
- P2) If some of S's cognitive states have an irreducible phenomenology, this would best explain S's immediate, noninferential knowledge of them; therefore (by inference to the best explanation),
- C1) Some of S's cognitive states have an irreducible phenomenology; and therefore,
- C2) There is such a thing as irreducible cognitive phenomenology.

It is not my current concern to evaluate the cogency of the Goldman-Pitt argument. I only wish to point out that it contains no phenomenological premises. Its first premise is epistemological and its second explanatory. Neither makes any claim about any phenomenal property actually being instantiated. The only propositions that make such a claim are the conclusions. Thus although the conclusions are phenomenological, no premise requires appeal to introspection. This represents, in effect, a fifth possible reaction to introspective disagreements: where such occur, they can be bypassed altogether by devising abductive arguments with no phenomenological premises. Call this the *abductive-bypass approach* to the problem of introspective disagreement.

Of the five possible reactions to introspective disagreement discussed here, I have avowed some sympathy for three: the introspective competence view, the deductive-bypass approach, and the abductive-bypass approach.⁵⁴ But the abductive-bypass approach is importantly superior in one respect, namely, the scope of putative phenomenal features it can target. Recall that the introspective competence view is most suitable for very general phenomenal features. A putative *sui generis* cognitive phenomenology would indeed be very general, but epistemic arguments along the above lines can target much more specific features as well.⁵⁵ Conversely, as I have pointed out elsewhere (Kriegel 2007), phenomenal features so general that they are present in every possible human experience (perhaps because they are *constitutive* of every experience) elude support by phenomenal contrast arguments. For such arguments require a contrast between two phenomenal episodes only one of which exhibits the feature at issue, but when a feature is universal (or constitutive) no such pair exists. In contrast, epistemic arguments for such features may still be mounted.⁵⁶

My preference for epistemic arguments in matters phenomenological will show itself at various places in this book. But I will develop my own version of this form of argument most fully in Chapter 3, where it will be used to argue that the attitude of entertaining a proposition exhibits its own proprietary phenomenology.

III. Looking Ahead

9. Plan of the Book

The purpose of the preceding has been to lay out the metaphysical and methodological foundations of a first-person inquiry into the scope and structure of the phenomenal realm. This book does not attempt to go very far in pursuit of this project. The core of the book consists in five studies of putative types of phenomenology that, antecedently, might qualify as second-layer phenomenal primitives: cognitive phenomenology, conative phenomenology, the phenomenology of entertaining, emotional phenomenology, and moral phenomenology.

Two other highly general putative types of phenomenology will be discussed in the Conclusion and Appendix: the phenomenology of imagination and the phenomenology of freedom.

To deprive myself of a punch line: I will end up avowing significant credence in cognitive, conative, imaginative, and entertaining phenomenology being second-layer phenomenal primitives. Emotional and moral phenomenology, by contrast, are reducible to combinations of other second-layer types of phenomenology, while the phenomenology of freedom is plausibly a lower-layer phenomenal determinable. Or so I will argue.

In addition to cognitive, conative, imaginative, and entertaining phenomenology, it is taken for granted in this book that perceptual phenomenology and algedonic phenomenology are second-layer primitives. This gives us a tentative list of six second-layer phenomenal primitives. My answer to Q is thus tentative but precise: six!



Chapter 1 is devoted to cognitive phenomenology. Although I support the Goldman-Pitt epistemic argument for cognitive phenomenology, my primary goal in that chapter is to devise a completely new type of argument for sui generis cognitive phenomenology. The argument relies on a thought-experiment in which the mental life of a person lacking any functioning sensory systems still exhibits indicators of phenomenality. A subsidiary goal of the chapter is to sketch an approach to the nonreductive characterization of such a primitive cognitive phenomenology.

Chapter 2 concerns conative phenomenology. The first part of the chapter offers an argument for primitivism about conative phenomenology, based on a series of phenomenal-contrast cases. The second part focuses on the *character* of conative phenomenology. In the functionalist literature that has dominated the philosophy of mind of the last generation or two, the paradigmatic conative state is considered to be desire. But while desire may be *functionally* paradigmatic, it is an open question which conative state is *phenomenologically* paradigmatic. I argue that the mental episode of *deciding and then trying to* ϕ is the fundamental form of conative phenomenology.

Chapter 3 is devoted to the act of entertaining a proposition. In the first part, I present a sustained epistemic argument for primitivism about the phenomenology of entertaining (which implies it is irreducible to cognitive phenomenology). In the second part, I offer a nonreductive characterization of entertaining in terms of its phenomenal feel and its connections to other propositional attitudes.

Chapter 4 focuses on emotional phenomenology. Part of the problem in discussing this topic is that the very nature of emotion is a controversial issue. I therefore start with a defense of a novel version of the “feeling theory” of

emotion, according to which emotions are essentially phenomenal. To do so, I develop an account of emotional phenomenology as involving crucially elements of cognitive and conative phenomenology. It is only once we appreciate emotional phenomenology's full intricacy, I contend, that we are in a position to see emotion as essentially phenomenal. I then consider whether emotional phenomenology is so rich as to involve an irreducible type of phenomenology. Although I reach no decisive conclusion, the weight of evidence seems to favor reductivism.

Chapter 5 is concerned with moral phenomenology, a topic that has garnered quite a bit of attention in recent metaethical research (Kriegel 2008). Here too, the task is complicated by the fact that there is no consensus on what moral mental states are to begin with. The bulk of this chapter is accordingly devoted to developing and defending a specific account of moral commitments, an account I call *dual-process cognitivist internalism*. With this account in place, I proceed to the question of whether the phenomenal feel of moral commitments might be primitive. I conclude that it probably is not: moral phenomenology plausibly reduces to a combination of cognitive, conative, and emotional phenomenology.⁵⁷

In the Conclusion, I address three questions. First, I argue that the phenomenology of imagination is a second-layer phenomenal primitive. Although this book is focused on nonsensory phenomenology, and imagination is typically sensory, I take this question up to "complete the picture" of second-layer primitives. Secondly, I revisit the question of taxonomy raised above. Finally, I briefly discuss a number of potential candidates for a seventh second-layer phenomenal primitive.

There is also an Appendix on the phenomenology of freedom, which plausibly is not a *second*-layer phenomenal determinable at all. My first-order goal is to articulate and defend a substantive characterization of the feeling of freedom. I defend three main theses, perhaps the most important of which is that freedom-experience involves a phenomenology as of compatibilist rather than libertarian freedom. The chapter's second-order goal is to explore some innovative methods for pursuing a first-person inquiry into phenomenal life. For this reason, much of the discussion is dedicated to methodological matters, which I suppose takes us full circle.

10. The Question of (Intellectual) Value

Let me close by addressing a pertinent question: *Who cares?* In other words, why should we expend intellectual energies on discovering the second-layer phenomenal primitives? Why does it matter? What hangs on it?

This is a question about the intellectual value of the project undertaken here. One way of answering the question is by pointing out research areas of

recognized intellectual value for which the present project may have implications. Consider research on the nature and structure of epistemic justification. Some philosophers have recently argued that perceptual experiences can confer positive epistemic status on perceptual beliefs purely in virtue of their phenomenal character (Huemer 2001, Pryor 2005). Thus, purely because of what it is like to have a perceptual experience as of a brown table, the experiencer is *prima facie* justified in believing that there is a brown table before her. Now, if perceptual phenomenology has this kind of justification-conferring power, then a *sui generis* cognitive phenomenology may as well—provided there is such a phenomenology. Perhaps purely because of what it is like to have a cognitive experience as of identity being a transitive relation, the experiencer is *prima facie* justified in believing that identity is a transitive relation (Chudnoff 2011a, Bengson forthcoming).⁵⁸

Similarly, the study of conative, emotional, and moral phenomenology may have implications for ethics and metaethics (Kriegel 2008). Thus, many ethicists have held that one is *prima facie* ethically justified in promoting pleasure and demoting pain (you need not be a consequentialist to hold this). It is, moreover, quite plausible to maintain that the reason it is good to demote pain has to do with the phenomenal character of pain—the unpleasant way it feels like to experience it. If so, the phenomenal character of pain and pleasure have the power to confer *prima facie* ethical justification on actions that might cause them. Now, if the algedonic phenomenology of pain and pleasure has this power, the same may hold of a *sui generis* emotional phenomenology involving subtler types of positive or negative affect—provided there is such phenomenology.⁵⁹ In this way, the intellectual value of the present project could be demonstrated by its implications for other areas of inquiry.

It may be worth insisting, however, that the project's intellectual value is not *only* instrumental in this way. The project also has *intrinsic, noninstrumental* intellectual value. There are reasons to train one's curiosity on the question of phenomenal primitives for its own sake, and regardless of its implications for other questions. Consider that early discussions of phenomenal consciousness in analytic philosophy of mind were concerned primarily with its problematic relation to the physical world, almost never with its internal variety and structure. Such discussions have tended to focus on simple, uncontroversial cases of phenomenal consciousness, most notably visual experiences (especially of red) and algedonic experiences (especially bodily pain). More recently, however, an a-reductive interest in phenomenal consciousness has started to gain traction. It has become apparent that many interesting philosophical questions arise already *within* the realm of the phenomenal, and not only regarding its relation to the physical realm. It has also become apparent that, thanks partly to the prolonged reign of exclusively reductive interest in phenomenal consciousness, our understanding of the internal variety and structure of the phenomena of

consciousness has remained limited. It is partly in this context that the debates over cognitive phenomenology, the phenomenology of agency, the richness of perceptual phenomenology, and the like have commanded increasing attention.

The project of searching for the phenomenal primitives is just a generalization from those debates, as is the yet more general project of mapping out the structure of the phenomenal realm. Insofar as there are facts of the matter targeted by these projects, it would certainly be desirable to *know* those facts. There are legitimate questions about whether we *can* know such facts, at least in the same sense in which we know other facts. But it is hard to deny that knowing such facts would be valuable—and valuable in and of itself, that is, intrinsically. After all, they are, more than any others, facts about *us*.⁶⁰