Phenomenal, normative, and other explanatory gaps: A general diagnosis

by Neil Mehta

Forthcoming in *Philosophy and Phenomenological Research*. Please cite the final published version.

ABSTRACT: I assume that there exists a general phenomenon, the phenomenon of the **explanatory gap**, surrounding consciousness, normativity, intentionality, and more. Explanatory gaps are often thought to foreclose reductive possibilities wherever they appear. In response, reductivists who grant the existence of these gaps have offered countless local solutions. But all such reductivist responses have had a serious shortcoming: because they appeal to essentially domain-specific features, they cannot be fully generalized, and in this sense these responses have been not just local but parochial. Here I do better. Taking for granted that the explanatory gap is a genuine phenomenon, I offer a fully general diagnosis that unifies these previously fragmented reductivist responses.

That we, like Descartes, can conceive of phenomenally unconscious physical duplicates of ourselves; that we, like Moore, can coherently ask whether what we desire to desire is truly good; that we, like Searle, can imagine an operator in a Chinese Room who lacks the slightest understanding of Chinese – I regard these facts, and many more, as mere instances of a general and philosophically pervasive phenomenon that I call the **explanatory gap**.[1] Explanatory gaps are often thought to foreclose reductive possibilities wherever they appear.

Here I set myself three tasks. First, I show that these familiar anti-reductivist arguments based on explanatory gaps can be fitted to a common template. On the one hand, we have standard reductive claims, like the claim that water reduces to (say) H_2O. Such reductive claims are entailed by an *a priori* conceptual truth stating a possible condition for *what it is*, essentially speaking, to be the entity in question, together with outside information of a certain limited kind telling us what meets that condition. In this sense, all standard reductive claims are *transparent*. On the other hand, we have gappy reductive claims – claims purporting to reduce the phenomenally conceived to the non-phenomenally conceived, the normatively conceived to the non-normatively conceived, and so on. What distinguishes these gappy reductions is that we are missing the relevant kind of *a priori* conceptual truths about essences.[2] As a consequence, such reductions are *never* transparent. The anti-reductivist would have us conclude, by inference to the best explanation, that the gappy reductions are spurious (§1-§2).

Many reductivists have responded by denying the existence of the asymmetry, and thereby denying that explanatory gaps are real. Though I, too,

---

[1] The term "explanatory gap" is of course most familiar from the phenomenal case and originates in Levine (1983).

[2] As I suggest in a later footnote, a similar argument template can be constructed using identity-based rather than essence-based models of reduction. Identity-based models appear in Lewis (1972), Jackson (1996), and Chalmers (2012).

am a reductivist, I do not follow these thinkers. I concede the reality of the explanatory gaps without argument.

Other reductivists have joined me in conceding the reality of the explanatory gaps, but unlike me, they offer *one-off* responses to the many explanatory gap arguments fitting the anti-reductivist's template. For example, according to the **phenomenal concept strategy**, phenomenal concepts have peculiar semantic features that generate the phenomenal explanatory gap – perhaps these concepts are distinctively recognitional, indexical, or quotational.[3] According to the **expressivist strategy**, meanwhile, normative concepts are distinctively plan-laden or motivationally efficacious, and these peculiar semantic features generate the normative explanatory gap.[4] And so on.

Each of these peculiar semantic features is meant to exempt the referent of any concept with that feature from the otherwise sound anti-reductivist conclusion. Troublingly, however, no semantic feature on any of these lists is shared by every concept in every domain associated with an explanatory gap. Phenomenal concepts are not usually thought to be plan-laden or motivationally efficacious; normative concepts are not usually thought to be recognitional, indexical, or quotational. As formulated, then, these diagnoses are *essentially* piecemeal: they rule out the very possibility of a general reductivist diagnosis.[5]

I aim to do better: my second task in this paper is to develop a fully general diagnosis of any explanatory gap (§3-§4). I say that every explanatory gap is the predictable symptom of a concept that is semantically **basic**. I therefore hold that any local semantic feature generates an explanatory gap only insofar as, and only because, a concept with that feature is basic. In this way, my diagnosis folds these previously fragmented strategies, as well as any yet-to-be-imagined strategies of the same type, into a single deep explanatory structure.[6]

---

[3] Defenders of various incarnations of the phenomenal concept strategy include Loar (1990/1997), Hill and McLaughlin (1999), Carruthers (2000), Tye (2000), Perry (2001), and Papineau (2002).

[4] See Gibbard (1990) and (2003) and Blackburn (1993) and (1998). Of course, expressivism has been thought attractive for certain other reasons as well. Arguably, for example, it explains the motivational significance of normative beliefs.

[5] Two versions of the phenomenal concept strategy are piecemeal, but not essentially so. First is the view, due to Hawthorne (2002) and Braddon-Mitchell (2003), that phenomenal concepts are **conditional concepts**, referring to certain non-physical properties if these exist and otherwise referring to physical properties. Though this diagnosis can perhaps be fully generalized, it is incompatible with my own. I reject it for reasons discussed in fn. 33.

A second generalizable version of the phenomenal concept strategy is due to Yetter-Chappell and Chappell (2013), who hold that phenomenal concepts are inferentially isolated from physical concepts and who extend the point to normative concepts. I believe that this diagnosis is right, insofar as it goes, but it remains to be seen whether this fact about inferential isolation is itself better explained in reductivist-friendly or anti-reductivist terms. Here I will argue for the former approach, as well as fully generalizing a strategy that Yetter-Chappell and Chappell see as covering only two cases.

[6] In Mehta (2013a), I develop this semantic diagnosis of the explanatory gap surrounding phenomenal consciousness. Here I fully generalize that solution, with updates as necessary along the way. The closest anticipation of my diagnosis appears in Levine (2010), who in passing makes the schematic point that basic phenomenal concepts suffice to explain the existence of the phenomenal explanatory gap. But besides not generalizing the point, Levine also explicitly rejects the existence of explanatory gaps, as characterized below, and hence rejects the diagnosis to be advanced here.

My third and final task is to show that this reductivist-friendly diagnosis is superior to the anti-reductivist alternative. I show that the anti-reductivist must accept my own diagnosis, while adding a further commitment: that every basic concept of a certain sort must refer to a metaphysically primitive entity. But this commitment, I argue, cannot do any explanatory work (§5). And while the anti-reductivist might suggest that we cannot confirm or disconfirm reductive hypotheses without appeal to explanatory gaps, I show that there is no mystery about how else to proceed (§6).

1. The profile of an explanatory gap

As the ambition of this paper is general, I must operate primarily at a schematic level; I cannot properly treat any of the myriad particular explanatory gaps. Still, it is worth mentioning – *very* briefly – a few examples of the phenomenon.

Attend to your current visual experience, focusing on what it is like for you to see the scene before you. Perhaps you see a shiny red apple and upon introspecting your experience you become aware of **phenomenal redness**. Now consider any theory reducing phenomenal redness to certain entities[7] picked out purely via non-phenomenal concepts, such as neural firing patterns, contents of perceptual representations of your environment, or external perceptibles with which you are perceptually acquainted.[8] Any such reduction of the phenomenally conceived with that conceived non-phenomenally will appear deeply dissatisfying. Even if some such reduction is correct, it is not clear *how* it could be correct. This is the **phenomenal explanatory gap**.

Next consider a normative property such as goodness. Reflect on a range of philosophical theories reducing goodness to entities picked out purely via non-normative concepts. Perhaps goodness is the greatest pleasure for the greatest number, or the property most conducive to the exercise of distinctively human capacities, or that property which would elicit the most approval under certain conditions that are specified using only non-normative concepts. Still, it apparently remains an open question whether something having the property in question is truly good.[9] The question continues to seem sensible regardless of what further non-normative background knowledge we might possess. This is the **normative explanatory gap**.

Turn next to intentional properties, such as the property of believing that snow is white. It may be thought that intentionality reduces to something conceived in purely non-intentional terms: perhaps to believe that snow is white

[7] I use the term "entity" as a maximally permissive sortal, covering states, processes, events, objects, properties, relations, etc.
[8] These are the mental paint, representationalist, and naïve realist theories of phenomenal consciousness. For defenses of standard versions of these theories, see respectively Block (1996); Dretske (1995), Tye (1995), and Lycan (1996); and Campbell (2002), Martin (2002) and (2006), Sturgeon (2008), and Fish (2009).
[9] Moore (1903). There have been many contemporary attempts, varying greatly in their faithfulness to Moore, to revive the open question argument. See Horgan and Timmons (1991), Rosati (1995), Altman (2004), Strandberg (2004), Nuccetelli and Seay (2007), Bedke (2012), Pigden (2012), and Sutton (2014).

is simply to behave or be disposed to behave in certain ways, or to have internal states suitably causally related to one another, or to be suitably biologically constituted. But it remains mysterious exactly how any such reduction could be true – how believing that snow is white could come to *nothing more than* meeting a non-intentionally conceived condition.[10] This is the **intentional explanatory gap**.

It is noteworthy that these explanatory gaps persist even when one attempts to reduce the entity in question to certain *metaphysically primitive* entities, as long as these primitives are considered purely via concepts drawn from outside the relevant domain. We do not solve the problem by reducing phenomenal redness, goodness, or intentionality to certain configurations of primitive goo, unless the goo is itself conceptualized as being phenomenal, normative, or intentional.[11]

Explanatory gaps abound in philosophy. There are, I suspect, similar explanatory gaps surrounding the self, space, time, causation, and more. But I will stick to discussing just two central instances, namely the phenomenal and normative explanatory gaps. Because these instances are superficially very different, it will be instructive to see that they share a deep structure, and it is a characterization of this structure rather than an inventory of its instances that is sought here.

We will need some terminology to frame the discussion. Think of a **concept** as the mental analog of a word, and stipulate that a concept must at least be a concrete mental representation belonging to a subject.[12] Then define a **conceptual family** as any group of concepts such that each concept in the group has suitably many conceptual connections to concepts within the group, but no concept outside the group has suitably many conceptual connections to concepts within the group. While this definition is imprecise, the intuitive idea is that a network of concepts amounts to a conceptual family in virtue of there being dense conceptual relations among the family members and sparse conceptual relations between family members and outsiders. We may leave open the possibility that a single concept may belong to multiple conceptual families, either nested or overlapping, as well as the possibility that a conceptual family may be as small as a single concept.

In the core instances of the explanatory gap phenomenon, it is obvious which conceptual family is most salient. For phenomenal concepts will plausibly form a conceptual family, given the relative plenitude of conceptual connections among phenomenal concepts and the relative paucity of conceptual connections

---

[10] I have mentioned behaviorism, functionalism, and the biological theory, respectively. Searle (1980) leverages the famous "Chinese room" thought experiment against functionalism, and the same argument obviously applies to behaviorism, as well. But Searle's own biological theory seems no less mysterious – what has the content of belief to do with *biology*?

[11] See for comparison Lewis (1988) on phenomenal consciousness and Moore (1903) on normativity.

[12] The term "concept" is sometimes used instead to refer to an abstract entity which is a part, perhaps improper, of a proposition. That will not be my usage here. This matter is merely terminological, as those who use the term "concept" in this alternative sense should also concede the existence of concepts in the sense of this paper: *something* about us must explain our representational capacities.

between phenomenal and non-phenomenal concepts. For similar reasons, normative concepts will plausibly form another conceptual family.

The two parties to the dispute are the **reductivist** and the **anti-reductivist**, so let us next understand what divides them. Their dispute concerns the referents of the concepts within some given gap-generating conceptual family F. The basic issue is that the reductivist accepts, and the anti-reductivist denies, that any entity considered via a concept in family F will be susceptible to an **interfamily reduction**. Let me explain.

I will work with an **essentialist model** of reduction.[13] According to this model, x **reduces** to y just in case the **essence** of x − *what it is* to be x − is to be y. Symbolically, I will represent the claim that x reduces to y as $\mathbf{x \leftarrow_E y}$: the subscript "E" stands for "essence," and the arrow pointing from y to x indicates that y is metaphysically prior to x.

I am thinking of essences in a broadly Aristotelian way. (Throughout, when I speak of the essence of an entity, I am always referring to its *complete* essence and never merely to some proper part thereof.) To recall a familiar example: necessarily, something is Socrates just in case it is a member of the set whose only member is Socrates. Even so, it is not the case that *what it is* to be Socrates is to be a member of the set whose only member is Socrates; that does not characterize the *essence* of Socrates.[14]

By using the essentialist model, we do justice both to the *asymmetry* and to the *hyperintensionality* of reduction. For suppose that water reduces to H_2O. Then according to the present account, what it is to be water − the essence of water − is to be H_2O. This treatment of reduction is asymmetric, for it is not true that what it is to be H_2O is to be water; presumably, what it is to be H_2O has something to do with protons, electrons, and the like. Our treatment is also hyperintensional. For if H_2O happens to be my favorite chemical, then it is necessarily true that H_2O is *actually* my favorite chemical − but it is not true that what it is to be water is to be my actual favorite chemical. Put symbolically: water $\leftarrow_E H_2O$, but $\sim(H_2O \leftarrow_E$ water), and $\sim($water $\leftarrow_E$ my actual favorite chemical).

Since I am treating reduction as hyperintensional, I will find it useful to say that a concept of some entity x is **essence-tracking** just in case the subject who possesses that concept thereby has a grasp, perhaps implicitly, of some possible condition for *what it is* (completely) to be x. We will see some examples of essence-tracking concepts below, but whenever I discuss some reductive claim of the form $x \leftarrow_E y$, I will normally assume that x and y are each considered via, and only via, their (complete) essences.[15]

It is now tempting to define anti-reductivism about some domain corresponding to conceptual family F as the rejection of *every* attempted reduction of any entity considered via an essence-tracking concept in F. Tempting, but wrong, since an anti-reductivist about a domain may accept reductions *within* the

---

[13] The thrust of what I have to say can be said even if we use an identity model of reduction. See fn. 24.

[14] I thus have in mind something like the understanding of essence introduced in Fine (1994a), and in particular what Fine (1994b) calls a constitutive essence. My model of reduction is inspired by Rosen (2010) and Fine (2015).

[15] I suspect that just about every concept is essence-tracking, but I will not rely on that idea here.

domain. An anti-reductivist about phenomenal consciousness might accept the reduction of phenomenal orangeness to some combination of phenomenal redness and phenomenal yellowness, and an anti-reductivist about normativity might likewise accept the reduction of what is good for a subject to what that subject has a certain kind of reason to desire.

Better, then, to define anti-reductivism as follows. Say that for a claim of the form "$x \leftarrow_E y$" to be an **interfamily** reductive claim is for x to be considered via some essence-tracking concept C belonging to conceptual family F and for y to be considered via its essence but solely via concepts outside family F. The claim is instead an **intrafamily** reductive claim if y is considered via at least one concept within family F. An **anti-reductivist** about some domain corresponding to some conceptual family F is then someone who rejects any *interfamily* reduction of any entity considered via any essence-tracking concept in F. By contrast, for every entity considered via an essence-tracking concept in F, a **reductivist** about the corresponding domain accepts some interfamily reduction of that entity.[16] Paradigms of the interfamily reductive claims in dispute between the reductivist and the anti-reductivist include the claims that phenomenal redness $\leftarrow_E$ such-and-such neural firing pattern and that goodness for a subject $\leftarrow_E$ what the subject desires to desire.

In this paper, I discuss only the anti-reductivist who wishes to defend her view by appeal to the explanatory gap phenomenon, as characterized below. I set aside without comment any anti-reductivist who would understand the explanatory gap phenomenon differently, as well as any anti-reductivist who does not appeal to explanatory gaps at all. Let my opponent present her case, then.[17] She proposes that there is an asymmetry between standard reductive claims (whether inter- or intrafamily) and gap-generating reductive claims. In particular, she suggests, any standard reductive claim is, in a certain epistemic sense, *transparent*, while no gap-generating reductive claim is transparent in this sense. The best explanation for the asymmetry is that the gap-generating reductive claims are all false. This suggestion requires some unpacking.

Begin with the definition of transparency. Consider some arbitrary entity x, considered via essence-tracking concept C belonging to family F. For the reductive claim that $x \leftarrow_E y$ to be **transparent** is for it to be the case that:

> (T1) Purely *a priori* conceptual information entails a possible condition for what it is to be x, and

---

[16] It may appear that this distinction between reductivism and anti-reductivism is not exhaustive – mightn't one think that some but not all entities considered via concepts in some family are susceptible to interfamily reductions? I think not: I would understand the notion of a conceptual family in such a way that this is impossible. As support for this idea, notice how unusual it is to hold that *some but not all* phenomenal entities can be reduced to entities conceived purely non-phenomenally, or that *some but not all* normative entities can be reduced to entities conceived purely non-normatively. Should that seem to be the case – e.g., if we want to distinguish genuinely normative reasons from merely institutional "reasons" – presumably we should say that the entities apparently susceptible to interfamily reduction are not actually phenomenal or normative after all.

[17] I do not intend the argument to be presented on the anti-reductivist's behalf to be mere exegesis of any actual anti-reductivist proposal, though it has close connections to many such proposals. Rather, I regard the argument as the best general version of an anti-reductivist argument based on the phenomenon of the explanatory gap.

(T2) full *a priori* or *a posteriori* information, *not considered via any concept from family F*, entails that some entity *y* meets those conditions.[18]

Let us take the reductive claim that water $\leftarrow_E$ H$_2$O as our paradigm of a transparent reductive claim.[19] Suppose that the ordinary concept of water belongs to a small family of concepts that includes the concept of the aquatic and the concept of hydration. Arguably, it is an *a priori* conceptual truth that *if* there is a single kind of entity *y* that actually plays the **water role** – e.g., that is transparent and odorless, that fills the rivers and the lakes, that causes me to have certain experiences, and so on – *then* water $\leftarrow_E$ *y*. If this is indeed an *a priori* conceptual truth, then the concept of water is an essence-tracking concept: the subject who possesses the concept of water thereby has a grasp of some possible condition for what it is to be water. So condition (T1) is satisfied.

Further, it is plausible that given full *a priori* or *a posteriori* information not considered via any concept in the water family, it follows that there is a single kind of entity, namely H$_2$O, that actually plays the water role. For start with all microphysical truths, considered via the concepts of completed physics. These will contain or entail information about the distribution and behavior of H$_2$O, including the information that H$_2$O is located in certain regions and behaves dynamically in certain ways. Then add all phenomenal truths, which will further entail that H$_2$O looks, tastes, and feels like water. Also add indexical truths identifying, for example, who and where I am. These will rule out the possibility that, say, water is XYZ rather than H$_2$O, as it will entail that the stuff that causes *me* to have experiences as of water, and that is distributed a certain way in *my* environment, is H$_2$O rather XYZ. Finally, add a totality truth to the effect that the world is a minimal world meeting the previous description. This will guarantee that H$_2$O is the *only* kind of entity that plays the water role. Yet none of the information required for this entailment needs to be considered via any concept in the family of water concepts. So condition (T2) is satisfied.

Thus, the anti-reductivist suggests, the claim that water $\leftarrow_E$ H$_2$O is transparent. For *a priori* conceptual information gives us a possible condition for what it is to be water, and full information considered without the use of any concept in the water family then entails that H$_2$O meets that condition.[20] Notice also that the claim that water $\leftarrow_E$ H$_2$O is an interfamily reductive claim. Thus, even the boldest and most systematic anti-reductivist should have no *general* opposition to interfamily reductive claims; she should oppose interfamily reductive claims *only in certain domains*.

---

[18] Strictly speaking, we should relativize this entailment to some conceptual family *F*, since we have left open the possibility that a concept may belong to multiple conceptual families. I leave this relativization implicit below.

[19] The claim that water $\leftarrow_E$ H$_2$O may be an oversimplification (see Leslie (2013)), but the oversimplification is harmless with respect to the basic point here. My subsequent treatment of this reductive claim is taken almost directly from Chalmers (1996) and (2012), Jackson (1998), and Chalmers and Jackson (2001); the key differences are made explicit in a later footnote. I also draw freely on my earlier discussion of the example in Mehta (2013a).

[20] But for skepticism that this claim is transparent, see Block and Stalnaker (1999), Byrne (1999), Levine (2001), Papineau (2011), Diaz-Leon (2011), and Schaffer (forthcoming).

For, she may argue, transparency is notably absent for all *interfamily* reductive claims in domains which generate explanatory gaps.

Return first to any attempt to reduce something conceived phenomenally to entities conceived purely non-phenomenally. All phenomenal concepts, the anti-reductivist may assume, form a family. But we do not have the kinds of *a priori* conceptual truths that we need to underwrite transparent interfamily reductions.

After all, certain phenomenal entities – such as phenomenal redness and greenness, perhaps –are not associated with any *a priori* conceptual truths at all stating possible conditions for what it is to be those entities. Other phenomenal entities, meanwhile, arguably are associated with *a priori* conceptual truths stating possible conditions for what it is to be those entities: perhaps it is an *a priori* conceptual truth that what it is to be phenomenal orangeness is to be the kind of entity that is related in certain ways to phenomenal redness and phenomenal yellowness. The trouble here, however, is that all of these *a priori* conceptual truths must themselves be considered using at least one other concept in the phenomenal family.

So no matter how much information we possess that is considered purely via concepts outside the phenomenal family, this information will not entail that some entity *y* meets the conditions laid out by these *a priori* conceptual truths. Either there will be no *a priori* conceptual truth of the right form, or there will be such a truth, but it will be possible to consider that truth only if we use other phenomenal concepts. *A priori* conceptual information alone never lets us break out of the phenomenal family. And that is why we find it perfectly conceivable that there might be worlds just like ours in all non-phenomenal respects, but in which our counterparts have no phenomenal experiences.[21],[22]

Or return to the normative domain. Any normative concept, the anti-reductivist may reasonably assume, belongs to a family that includes both "thin" normative concepts, such as the concept of goodness and the concept of a reason, as well as "thick" normative concepts, such as the concepts of bravery and generosity. Some normatively conceived entities – certain reasons, perhaps – are not associated with any *a priori* conceptual truths stating possible conditions for what it is to be those entities. All other normatively conceived entities, meanwhile, are associated with *a priori* conceptual truths of the right kind. Perhaps, for example, it is an *a priori* conceptual truth that for something to be good for a subject is for there to be a certain kind of reason for her to desire it. But every such truth can be considered only if we use at least one concept in the normative family.

[21] It may appear that some relevant truth about reference, such as the (putative) truth that the concept of phenomenal redness refers to such-and-such neural firing pattern, permits such an entailment. But this appearance is misleading. For reflect on how the concept of phenomenal redness is to be considered in a thought entertaining this claim. If the concept is considered as the concept *of phenomenal redness*, then considering this claim will use the very concept of phenomenal redness, which is not permitted. If on the other hand the concept is considered in some way that does not require the use of phenomenal concepts, then consideration of this truth is permitted but will not entail anything interesting about phenomenal redness conceived as such, and hence will not entail that phenomenal redness (conceived as such) $\leftarrow_E$ such-and-such neural firing pattern.

[22] Some would defend the transparency of interfamily reductive claims even in the phenomenal case. See Lewis (1966), Lormand (2004), Jackson (2007), Fish (2008), and Sundström (2017). I believe that McGinn (1991) would also defend transparency in this case, though he doubts that human beings are equipped to understand the relevant entailment.

So no matter how much information we have considered via concepts *outside* the normative family, this information will not entail that some entity *y* meets the conditions laid out by these *a priori* conceptual truths about normative essences. That is why we find it sensible to ask, "This is what I desire to desire, but is it truly good?"

The normative case does present one important complication: there are arguably many *a priori* conceptual truths linking normativity to conditions that can be considered via purely non-normative concepts. For example, it is arguably an *a priori* conceptual truth that if an act involves killing a person purely for sport, then that act is morally wrong. Such truths might seem superficially similar to the essentialist *a priori* conceptual truths required for a transparent reduction.

But there is a profound difference: these normative truths, rather than stating possible conditions for *what it is* to be (e.g.) wrong, just state *metaphysically necessary and/or sufficient* conditions for the property of being morally wrong to be *instantiated*. That is not enough to underwrite a transparent reduction, since we can know that moral wrongness is instantiated without knowing what moral wrongness *is*.[23]

This difference explains why we continue to find it mysterious how it could be that, e.g., goodness $\leftarrow_E$ what I desire to desire, *even if* we become convinced that goodness and what I desire to desire have functional profiles that are identical in all non-normative respects. For we cannot rule out the possibility that (i) necessarily, something is good just in case the subject desires to desire it, but (ii) being good and being what I desire to desire are distinct properties.[24]

Thus, the differences between the phenomenal and normative explanatory gaps are merely superficial. For in both domains, and arguably in many other domains as well, we must confront a difficult question: given that *all* typical reductive claims are transparent, why are there *no* transparent interfamily reductive claims in these gap-generating domains? The best explanation, according to the anti-reductivist, is that all gap-generating reductive claims – indeed, all reductive claims that are not transparent – are simply false.[25]

---

[23] If there are *a priori* conceptual truths stating possible conditions, graspable via purely non-normative concepts, for *what it is* to be good, *what it is* to be a reason, etc., then there simply is no normative explanatory gap, and all appearances to the contrary are due to our lack of appreciation of these *a priori* conceptual truths.

[24] This point goes through even if we accept an identity-based model of reduction. The point would then be that an *a priori* conceptual truth might yield a non-normatively conceived condition under which a normative property is *instantiated*, but it will not yield a non-normatively conceived condition under which a property is *identical to* a normative property.

Thus I oppose those who take explanatory gaps to support anti-reductivism in the phenomenal case, but not in the normative case (see Chalmers and Jackson (op. cit.)). I also regard this point as the normative anti-reductivist's best response to reductive normative functionalism, as developed by Jackson and Pettit (1995, p. 28); perhaps the point also targets Smith (1994) as well, though I am uncertain whether his normative functionalism is in fact *reductive*.

[25] It is sometimes said in response that an explanatory gap arises in any Frege case, and so the explanatory gap argument cannot reasonably be taken to undermine gappy reductive claims. (See the response of Schroeder (2007) to the open question argument, for example; the nuanced conclusions of Kalderon (2004) are also in this vicinity.) Arguably, however, the explanatory gap argument as formulated here avoids such concerns. Compare Chalmers and Jackson (2001, pp. 354-356).

In sum, wherever the anti-reductivist finds an explanatory gap, she can wield an argument fitting this template:

(EGA1)   Every standard reductive claim is transparent.

(EGA2)   For any entity x, considered via some essence-tracking concept C in family F, no interfamily reductive claim of the form $x \leftarrow_E y$ is transparent.

(EGA-C)   Therefore, every such interfamily reductive claim is false.[26]

The conclusion is intended as an inference to the best explanation from the asymmetry identified in premises (EGA1) and (EGA2).

2. Clarifications of the argument; possible responses

It is important to note that we do not have an explanatory gap every time that an entity fails to figure in any transparent interfamily reductive claim. Suppose, for example, that the concept of a quark is much like the concept of water. In particular, suppose that it is an *a priori* conceptual truth that *if* there is a single physical kind of entity that plays a certain theoretical role – the **quark-role** – *then* what it is to be a quark is to be an entity belonging to that kind. Now, it might just turn out that quarks are metaphysically primitive. That is, it might just turn out there is a single physical kind of entity that plays the quark role, but that kind of entity is just a *quark*. Then any attempt to reduce quarks transparently – indeed, any attempt to reduce quarks at all – is bound to fail.

But this would not be an instance of an explanatory gap, since the failure would be purely *a posteriori*. In these kinds of cases, the first stage required for a transparent interfamily reduction is successful: there is an *a priori* conceptual truth stating a possible condition for what it is to be the entity in question, where that condition can be considered using only concepts outside family F. The reduction fails *only at the second stage* – the stage at which we gather information considered via concepts outside the relevant family – for it turns out that there is no entity meeting the possible condition that we gathered in the first stage. But it was perfectly *a priori* possible that the second stage would be successful; it just so happened that it was not.

What is distinctive about a gap-generating entity is that attempted interfamily reductions of it fail to be transparent *at the first stage – the a priori conceptual stage*. In particular, with a gap-generating entity x considered via concept C in family F, there is no *a priori* conceptual truth stating a possible condition for what it is to be x, where that condition can be considered using only

---

[26] Note that (EGA1) should not be replaced with the claim that every *true* reductive claim is transparent. For while the argument would then be deductively valid, it would also beg the question against the reductivist, who holds that many true reductive claims are not transparent. Levine (2010) and (2014) notes the circularity of an argument concerning phenomenal consciousness fitting this template and attributes the circular argument to Chalmers and Jackson (2001). Setting aside questions of Chalmers-Jackson exegesis, I agree with him about the circularity of the argument he discusses and hence believe that the argument template in the text is much more dialectically effective.

concepts outside family *F*. This will occur either (i) because there are no *a priori* conceptual truths stating possible conditions for what it is to be an *x* at all, or (ii) because there are such truths, but each of them can be considered only via some concept in family *F*. That is what it is for an entity to generate an **explanatory gap**. This definition is purely stipulative; the term "explanatory gap" has been given other definitions which should not be confused with this one.[27]

How can the reductivist resist the anti-reductivist's explanatory gap argument? One option is to reject one of the argument's two premises. If the reductivist rejects (EGA1) while granting (EGA2), then she says that even some standard reductions are not transparent. If she rejects (EGA2) while granting (EGA1), meanwhile, then she claims that there are transparent reductions not only in standard domains, but also in the allegedly gappy domains.[28]

Reductivists who offer these two responses endorse very different models of reduction. Nonetheless, they agree that the anti-reductivist simply fails in her attempt to identify a principled asymmetry between standard reductive claims and gappy reductive claims. In other words, they deny that the very existence of the explanatory gap phenomenon, as I am using that term. In my view, however, the explanatory gap phenomenon is genuine; like anti-reductivist opponent, I accept (EGA1) and (EGA2). I make no attempt to defend that approach here but simply take it as a working assumption.

Thus I am left with only one response: I must contest the inference to the anti-reductivist conclusion. The anti-reductivist says that her conclusion best explains the explanatory gap phenomenon. I say that the phenomenon admits of a better explanation – in particular, a purely semantic explanation.

As I have already stated, whereas previous versions of this response have been essentially piecemeal, mine will be fully general. I first introduce a neutral semantic framework (§3), from which I extract a purely semantic, reductivist-friendly diagnosis of the explanatory gap phenomenon (§4). Although the anti-reductivist will be tempted to supplement this semantic diagnosis with a further metaphysical one, I show that we should resist this temptation (§5-§6).

---

[27] The anti-reductivist faces one minor complication having to do with indexical concepts – concepts expressed with words like "I," "now," and "this." Suppose that these form a conceptual family, or at least a small cluster of conceptual families (the personal-indexical family, the temporal-indexical family, the demonstrative-indexical family, etc.). It is a familiar point from Perry (1979) and Lewis (1979) that any such family of concepts will generate an explanatory gap: there is no *a priori* conceptual truth stating a possible condition on what it is to be *me*, or to be *this* (demonstrated object), in a way that we can conceive without using any concept in the relevant family. But in at least some of these cases, we feel no resistance to interfamily reductions – certainly not in the case of ordinary material objects picked out via demonstrative concepts, for example.

It is standard for the anti-reductivist to concede that *in these special cases*, the explanatory gap argument template fails: there is an explanatory gap, but interfamily reductions may still be had. Of course, if the explanatory gap argument template is to be safely used in *other* cases, then the anti-reductivist owes a principled distinction between indexical cases and phenomenal, normative, and other cases. However, since I see many potential ways that the anti-reductivist might meet these requirements, I propose to simply concede that there is some such account to be had without fussing over its details. See Chalmers (2004, pp. 186-7) for one possible approach.

[28] She should not reject *both* premises; that response is plainly mad. It entails that phenomenal and normative reductive claims are transparent, while reductive claims about water are not. See previous footnotes for many examples of those who would reject either (EGA1) or (EGA2) in the phenomenal and normative domains.

3. A minimal semantic framework

I will in this section develop a **minimal semantic framework** that we can use to understand the explanatory gap phenomenon. I call the semantic framework *minimal* because, as I will argue, it should be accepted by any theorist, whether she is a reductivist or an anti-reductivist, and indeed whether she accepts or rejects the existence of explanatory gaps. The core idea is to divide all concepts into *basic* and *non-basic* concepts. Roughly speaking, for a concept to be non-basic is for it to link to its referent *by* first linking to one or more other concepts; in this sense a non-basic concept is semantically dependent. Roughly speaking, for a concept to be basic is for it to link to its referent without conceptual intermediary; in this sense a basic concept is semantically independent.

Let us make all of this precise. Say that for some concept *C* to **semantically depend** on concepts $C_1 \ldots C_n$ at time *t* is for the following two elements to ground the reference of concept *C*: first, a certain privileged psychological relation between concept *C* and concepts $C_1 \ldots C_n$ at time *t*, and second, the extensions of $C_1 \ldots C_n$ at time *t*. So as to leave the present framework compatible with the widest possible range of semantic theories, I will remain neutral about the precise character of the relevant privileged psychological relation. But perhaps the concept of a bachelor semantically depends in this way on the concepts of being a man and of being unmarried.

I intend for the notion of grounding used to characterize semantic dependence to denote a metaphysical "in-virtue-of" relation. I therefore make two further assumptions.

The first assumption imposes a "no-junk" condition on semantic dependence. To return to our example, from the fact that the concept of a bachelor semantically depends on the concepts of being unmarried and being a man, it should not follow that the concept of a bachelor semantically depends on the concepts of being unmarried, being a man, and being a lizard – the concept of being a lizard is junk. To be precise, the first assumption is that semantic dependence is **non-monotonic**: if *C* semantically depends on concepts $C_1 \ldots C_n$ at time *t*, it does not follow that at time *t* that *C* semantically depends on the concepts in any group which includes $C_1 \ldots C_n$.

The second assumption imposes a "no-circles" condition on semantic dependence. The idea is that if we start from some concept *C* and trace the chain of semantic dependence, we will never arrive back at *C*. To be precise, the second assumption is that semantic dependence is **group asymmetric**: if concept *C* semantically depends on concepts $C_1 \ldots C_n$ at time *t*, then at time *t* none of the concepts $C_1 \ldots C_n$ may semantically depend on any group of concepts which includes *C*; nor may any concepts on which $C_1 \ldots C_n$ semantically depend themselves semantically depend on any group of concepts which includes *C*; and so on. It follows from the second assumption that no concept semantically depends on itself.

We may now partition all concepts according to whether or not they stand in semantic dependence relations. For a concept to be **non-basic** at a certain time is for it to semantically depend on at least one of the subject's other concepts at

that time, while for a concept to be **basic** at a certain time is for it not to semantically depend on any of the subject's other concepts at that time. In a metaphor: the reference of a non-basic concept flows through other concepts to the world, while the reference of a basic concept flows directly to the world without conceptual intermediary. Bear in mind, though, that a concept's non-reference-determining relations to other concepts are not germane to its status as basic or non-basic.

The minimal semantic framework entails an important consequence: any subject must possess some basic concepts, at least if her stock of concepts is finite. Proof: pick any subject with finitely many concepts and list all of her concepts. Then iterate the following procedure: ask whether the concept at the top of the list semantically depends on any other concepts. If it does not, then stop. If it does, then remove it from the list, pick any concept on which it semantically depends, and move that concept to the top of the list.

Since the list is finite, the procedure is sure to halt. And when it does the concept that tops the list must be a basic concept: that concept cannot semantically depend on any concept(s) remaining on the list, since then the procedure would not have halted, nor can it depend on any concept(s) previously removed from the list, since semantic dependence is group asymmetric. This result should come as no surprise. Any genuine referential link must ultimately make contact with the world. It cannot be conceptual mediation all the way down.

I turn now to a defense of the minimal semantic framework. The framework rests on just two assumptions about semantic dependence: a "no-junk" condition and a "no-circles" condition. These are deeply plausible assumptions about grounding in general, and hence about semantic dependence in particular. The assumptions might fail to hold in a few exotic cases – might there be a God who metaphysically depends on Herself? – but semantic dependence is not an exotic case.[29,30]

One might reject the minimal semantic framework because one misunderstands what follows from it:

1. One might think that all concepts are basic. Yet apparently the minimal semantic framework says otherwise.

Not so. The minimal semantic framework entails that at least some concepts (of any subject with finitely many concepts) are basic. It does not entail that any concepts are non-basic.

2. One might think that there are no stable definitions. Yet if any concepts are non-basic, then apparently the non-basic concepts must have stable definitions.

Not so. For one thing, perhaps a single non-basic concept might semantically depend on different concepts at different times, while maintaining the same referent throughout. For another, perhaps a single concept might be

---

[29] For sympathetic discussion of the possibility of circles of grounding, see Bliss (2014).

[30] And even if the assumptions are rejected, the theorist who acknowledges the existence of explanatory gaps will need to accept some semantic framework in the vicinity to account for the existence of the *a priori* conceptual truths about essences to which she is committed. In particular, this alternative framework, like the one I have presented, must take on a semantic commitment to conceptual links and (as will be discussed below) an epistemic commitment to the *apriority* of those links. I therefore suspect that an anti-reductivist who champions any plausible alternative framework of this sort will encounter objections analogous to the ones I will raise below. I will not attempt to vindicate this suspicion here.

non-basic at one time and basic at another time – again, while maintaining its referent throughout. And even if we focus on a non-basic concept at a single time, we need not think of the concepts on which it semantically depends as *defining* it, in any traditional sense.

3. One might think that there is only a single reference relation R. Yet if we admit both basic and non-basic concepts, then apparently we must also admit at least two distinct reference relations: one between basic concepts and their referents, and another between non-basic concepts and their referents.

Not so. There might be only a single reference relation R even in this case. It just must be that *what makes it the case* that a concept bears R to its referent depends on whether or not the concept is basic. If the concept is non-basic, then it must bear R to its referent *because it stands in certain relations to other concepts* (which bear R to various referents of their own); while if the concept is basic, then it must bear R to its referent *for other reasons.*

4. One might think that two or more concepts can have their reference fixed jointly. Indeed, one might accept the **total holist** view that *all* of our concepts have their reference fixed jointly. Yet these views apparently violate the second assumption of the minimal semantic framework, the assumption of group asymmetry.

Not so! For suppose that certain concepts $C_1…C_n$ have their reference fixed jointly. If we are to maintain the second assumption, then we cannot say, e.g., that C_1 semantically depends on $C_2…C_n$ and that C_2 also semantically depends on C_1, $C_3,…C_n$. But we may instead say that the reference of each of these concepts is determined by a certain privileged relation between that concept and the world, on the one hand, and a certain privileged relation between that concept and all of the subject's other concepts, on the other hand. Interestingly, on this approach, none of the concepts $C_1…C_n$ has its reference determined solely by a privileged psychological relation to any group of the subject's other concepts along with the extensions of those concepts. A certain relation between each of these concepts and the world is also required to fix its reference. As a consequence, *all* of these concepts end up being basic, and total holism is just one version of the view that all concepts are basic. But all of this remains compatible with our framework.

The minimal semantic framework truly is minimal: its assumptions rest on deeply plausible metaphysical principles without unduly restricting the range of permissible semantic theories. I therefore urge reductivists and anti-reductivist alike to accept it.

Let us see what follows.

4. A semantic diagnosis of any explanatory gap

The minimal semantic framework gives us enough traction to state the precise conditions under which an explanatory gap will arise. In particular, I will argue, an explanatory gap arises just in case there is an essence-tracking basic concept within the relevant conceptual family.

We will need two tools before we make our diagnosis. First, we must note that any theorist who acknowledges the reality of explanatory gaps should accept

that semantic dependence relations underwrite the *a priority* of conceptual truths. E.g., those who acknowledge the reality of explanatory gaps should accept that if the concept of a bachelor semantically depends in a certain way on the concepts of being unmarried and being a man, then it is a conceptual truth that any bachelor is an unmarried man, and that conceptual truth is *a priori* because of these semantic dependence relations. For theorists who acknowledge explanatory gaps hold that there are many *a priori* conceptual truths about essences. Yet it is hard to see how that could be unless semantic dependence relations, which underwrite conceptual truths, also make those truths *a priori*.

The second tool that we will need is the notion of a **semantic dependence chain**. Say that *C* stands in a semantic dependence chain to some group of concepts *G* just in case *C* semantically depends on *G**, where we can get from *G* to *G** by iterating the procedure, as many times as we like, of replacing one or more of the concepts in *G* with any concept that semantically depends on the replaced concept(s). To take a toy example, suppose that the concept of a bachelor semantically depends on the concepts of being unmarried and of being a man, and suppose further that the concept of being unmarried semantically depends on the concept of negation and the concept of being married. Then the concept of a bachelor stands in a semantic dependence chain to the concepts of negation, of being married, and of being a man.

With these two tools in hand, we can determine exactly when an explanatory gap will arise. Recall that an entity *x*, considered via some essence-tracking concept *C* in family *F*, generates an **explanatory gap** just in case there is no *a priori* conceptual truth, considered solely via concepts outside *F*, stating a possible condition for what it is to be *x*. So consider an arbitrary entity *x*, considered via some essence-tracking concept *C* in family *F*, that we hope to reduce. There are three possibilities.

First possibility: *C* is a basic concept.

If this first possibility obtains, then *x* will generate an explanatory gap. For if *C* is a basic concept, then there will be no concepts which determine the reference of *C*, and hence there will be no *a priori* conceptual truth stating a possible condition for what it is to be *x*. If the concept of phenomenal redness is essence-tracking and basic, for example, then no reductive claim of the form "phenomenal redness $\leftarrow_E y$" will obey the standard model.

One interesting consequence follows immediately: if explanatory gaps are real, then there must be some non-basic concepts. For standard reductive claims are *not* supposed to generate explanatory gaps, yet every basic concept will generate an explanatory gap. So it must be that for any standard reductive claim, the reduced entity is being considered via a non-basic concept.

Second possibility: *C* is a non-basic concept, and *C* stands in a semantic dependence chain to some group of concepts *G* that includes *no* concepts from family *F*.

If the second possibility obtains, then *x* will not generate an explanatory gap. For there will be an *a priori* conceptual truth stating a possible condition for what it is to be *x*, where this condition can be considered purely via the concepts in *G* – which, we have stipulated, are all outside family *F*. That is precisely what is required for there not to be an explanatory gap.

To return to our stock example, suppose that the concept of water belongs to the family of water concepts and stands in a semantic dependence chain to a group of concepts none of which belong to this family. The latter concepts might include the concepts used to consider the entity that actually plays the water role – the entity that is transparent and odorless, that fills the rivers and the lakes, that causes me to have certain experiences, and so on.[31] Then it will be an *a priori* conceptual truth that (e.g.) if there is a single natural kind of entity y that actually plays the water role, then water $\leftarrow_E y$. This possible condition for what it is to be water can moreover be considered solely via concepts outside the family of water concepts.

I repeat that even if this second possibility obtains, there is still no *guarantee* that we will find a transparent interfamily reduction. For it is *a priori* possible that water is metaphysically primitive – in which case the single natural kind of entity that actually plays the water role would just turn out to be *water*. As I suggested earlier, certain cases might actually be like this: it might be that there is no explanatory gap surrounding quarks, but that quarks might nevertheless be metaphysically primitive. But if interfamily reductions do fail to be transparent in this way, then that failure will occur for purely *a posteriori* reasons having nothing to do with the explanatory gap phenomenon. Thus, if the second possibility obtains, then *whether or not* there actually turns out to be a transparent interfamily reduction, there is still no explanatory gap.

Third possibility: C is a non-basic concept, and any group of concepts to which C stands in a semantic dependence chain includes *at least one* concept from family F.

If the third possibility obtains, then x will generate an explanatory gap. To be sure, there will be at least one possible condition for what it is to be x. For C stands in a semantic dependence chain to at least one group of concepts, and that fact will underwrite an *a priori* conceptual truth stating a possible condition for what it is to be x. However, in this third possibility, any group of concepts to which C stands in a semantic dependence chain includes *at least one* concept in family F. So all of these *a priori* possible conditions for what it is to be x must themselves be considered via at least one concept in family F. Thus, we might find a transparent *intra*family claim of the form $x \leftarrow_E y$, but there will be no route to a transparent *inter*family claim of that form. Moreover, the route will be blocked for purely *a priori* conceptual reasons. And that constitutes an explanatory gap.

To return to some familiar possible examples: consider phenomenal orangeness, which perhaps can be transparently reduced to a certain combination of phenomenal redness and phenomenal yellowness, but which cannot be transparently reduced to any entities conceived purely non-phenomenally. Or consider goodness, which perhaps can be transparently reduced to what there is a certain kind of reason to desire, but which cannot be transparently reduced to anything conceived purely non-normatively.

Note that if this third possibility obtains – if C is non-basic and any group of concepts to which C stands in a semantic dependence chain includes *at least one*

---

[31] Or perhaps many of these concepts – the concepts of water, rivers, lakes, etc. – *together* stand in a semantic dependence chain to some distinct group of concepts. Then perhaps the concepts in that group help to *simultaneously* determine the reference of the concepts of water, rivers, lakes, etc.

concept from family F – then it must be that some concept within family F is basic. For if the chain of semantic dependence beginning with concept C were to bottom out in basic concepts that were all outside family F, then we would find ourselves in the second possibility rather than the third. If C is the concept of phenomenal orangeness, then these basic concepts might include the concepts of phenomenal redness and phenomenal yellowness; if C is the concept of goodness, then these basic concepts might include the concept of a reason.

The minimal semantic framework therefore yields a precise diagnosis of the explanatory gap:

> The **semantic diagnosis**: Any entity x, considered via concept C in family F, generates an explanatory gap just in case there is some essence-tracking basic concept in family F.

If the basic concept is C itself, as in the first possibility, then there can be no transparent reduction of x at all. If the basic concept is some concept other than C in family F, as in the third possibility, then there might be a transparent *intra*family reduction of x, but there cannot be a transparent *inter*family reduction of x.

The semantic diagnosis is *purely* semantic. It is therefore reductivist-friendly, as it has no metaphysical import whatsoever. And, I have argued, any theorist who accepts the reality of explanatory gaps should accept this diagnosis. Where does that leave the anti-reductivist?

5. Anti-reductivist diagnoses

The anti-reductivist, too, should accept this diagnosis – with a supplement. For she holds that any true reductive claim is transparent, and we have just seen that the referent of an essence-tracking basic concept can never be transparently reduced. Thus, the anti-reductivist is committed to the

> **Semantics-metaphysics link**: Any essence-tracking basic concept refers to a metaphysical primitive.[32]

In this section I evaluate this principle.

One immediate worry is that the semantics-metaphysics link rules out the possibility of phenomenal (or normative, intentional, …) **gunk**, i.e., a bottomless chain of phenomenal entities such that each entity reduces to phenomenal entities further down in the chain, but no entity in the chain reduces to anything considered via purely non-phenomenal concepts. Or at least the semantics-metaphysics link rules out the possibility that *our* basic phenomenal concepts refer to any entity in such a bottomless chain. Yet even for the anti-reductivist this conclusion seems too strong: she wants to rule out certain kinds of *interfamily* reductions, but she may be perfectly content to allow for such (endless) *intrafamily* reductions.

---

[32] Except for indexical concepts – see fn. 27.

Still, some anti-reductivists may be willing to bite this bullet, and in any case we will eventually encounter an anti-reductivist proposal that avoids the worry entirely. So I will table this concern for now.

Instead, I will press the following question: since the anti-reductivist is already committed to a semantic diagnosis of the explanatory gap phenomenon, what work is left for the semantics-metaphysics link to do?

5.1. A conjoined explanation

The simplest suggestion for the anti-reductivist is that the semantics-metaphysics link, *together with* the semantic diagnosis, is what explains the explanatory gap phenomenon. In other words, the suggestion is simply to conjoin the semantics-metaphysics link to the semantic diagnosis.

This suggestion can be briskly dismissed. For the semantics-metaphysics link is not explanatorily necessary. The purely semantic diagnosis, which is common to my account and the anti-reductivist's, *already* explains the explanatory gap phenomenon in full. The semantics-metaphysics link is not even explanatorily *helpful* – it adds nothing of relevance to the explanation. But it is ontically *pernicious*: it forces the anti-reductivist to treat certain properties, including some phenomenal and normative properties, as primitive, whereas the reductivist would identify these with properties countenanced by all. Accepting the semantics-metaphysics link, which is explanatorily idle but ontically inflationary, therefore leaves us with the worst of both worlds.[33]

5.2. Explaining the existence of essence-tracking basic concepts

But there is a better suggestion that the anti-reductivist might offer. She might agree with me that the semantic diagnosis by itself explains the explanatory gap phenomenon, while adding that the relevant semantic facts are *themselves* best explained in metaphysical terms. And that explanation might entail the semantics-metaphysics link. In other words, in response to the question, "Why are there explanatory gaps in certain domains?" the anti-reductivist now gives the same answer that I give: "Because there are essence-tracking basic concepts in each of the associated conceptual families." But the anti-reductivist sees a further question that deserves an answer: "Why do *those* conceptual families – e.g., the phenomenal,

---

[33] It is for a similar reason that I hold that the explanatory gap phenomenon by itself lends no support to the alternative reductivist-friendly **conditional concept diagnosis**: the diagnosis that an explanatory gap is the symptom of a conditional concept (see fn. 5). To put tersely a point that deserves more discussion: a conditional concept cannot be basic, as its reference semantically depends on the concepts required to consider the relevant conditional. Hence any friend of the conditional concept diagnosis who accepts the assumptions of this paper must admit that essence-tracking basic concepts *also* generate explanatory gaps. But, with this full account of the explanatory gap phenomenon already in hand, the conditional concept diagnosis is of no further help here. It must be defended on grounds independent of the explanatory gap phenomenon.

normative, and intentional families – include essence-tracking basic concepts?" So far I have given no answer to this question; the anti-reductivist wants to answer, "Because the essence-tracking basic concepts in those families all refer to metaphysical primitives."[34]

This anti-reductivist suggestion is much more promising than the last. It fills a genuine explanatory need, and in smart fashion. Before we evaluate it, however, let us get on the table an alternative, reductivist-friendly explanation for why we have the essence-tracking basic concepts that we do.

I propose an explanation that is built around this core idea: our minds include many specialized *non-conceptual* systems each dedicated to a particular evolutionarily important task, as well as a centralized *conceptual* system for more general-purpose processing. (One might or might not want to understand the distinction between these two types of systems in terms of the distinction between modular and central processing.) For each specialized system, we will find a corresponding family of concepts in the central system – concepts that are at the interface between the two systems, so to speak.[35]

Note well: I do not claim that every concept in such a family must be at the interface between the two systems, since we might use those concepts which are at the interface to form new concepts that are still in the family but that are no longer at the interface. Perhaps "thick" normative concepts, like the concepts of bravery and generosity, are like this. The idea is just that for every specialized system that interfaces with the central system, there will be a corresponding family of concepts.

I further hypothesize that each conceptual family of this kind will include some essence-tracking basic concepts. In particular, I hypothesize that the essence-tracking basic concepts within the family will be those that happened to develop within the central system to allow it to exchange information with the specialized system. To predict which conceptual families will include essence-tracking basic concepts, we should therefore look to the specialized systems that we happened to evolved. Here are some plausible candidates.

First, we plausibly have a specialized *perceptual system* to detect various environmental and bodily features; correspondingly, we have a family of phenomenal concepts. Second, we plausibly have a specialized *preference-governing system* to propel us to action – to propel us to seek mates and ripe fruit and to avoid predators, for example; correspondingly, we have a family of normative concepts. Third, we plausibly have a specialized *mind-reading system* that ascribes psychological attitudes to others (and to ourselves), so that we can predict and respond to the behavior of others (and ourselves); correspondingly, we have a family of intentional concepts.[36]

We now have two rival explanations for why we have the essence-tracking basic concepts that we do. The anti-reductivist explanation says that we have these essence-tracking basic concepts because of certain metaphysical facts, as encoded in the semantics-metaphysics link. The reductivist-friendly explanation says that we have these essence-tracking basic concepts because of contingent

---

[34] I thank [name removed], [name removed], and an anonymous referee for bringing this anti-reductivist idea to my attention.
[35] This proposal is inspired by Lormand (ms).
[36] Lormand (ms) develops these ideas in some detail.

facts about human psychology, as conditioned by evolutionary history. Which explanation is better?

I see one clear advantage for the anti-reductivist explanation: it is more elegant. The anti-reductivist says that every essence-tracking basic concept corresponds to a metaphysical primitive. What a pretty picture she paints! The reductivist-friendly picture is pleasing, but not as pleasing as this one.

But, I suggest, the anti-reductivist hypothesis does not score so well on other tests. Start with the test of *independent plausibility*: how plausible is each hypothesis independently of its ability to explain why we have the basic concepts that we do? The reductivist-friendly hypothesis has much independent plausibility, as this hypothesis rests on claims about psychology and evolutionary theory that have received ample empirical confirmation. The precise details of the proposal may need adjustment, but there are many ways of doing so without making any compromises in spirit.

But compare the semantics-metaphysics link, which is a central commitment of the anti-reductivist hypothesis. According to the semantics-metaphysics link, any essence-tracking basic concept must refer to a metaphysical primitive. At first blush, this idea might seem independently plausible, given that any metaphysically non-primitive entity must reduce to other entities. Suppose, then, that I form some essence-tracking concept C that refers to a metaphysical non-primitive. Arguably, since C is essence-tracking, it would have to track the reduced entity *by* tracking the reducing entities – in which case it will semantically depend on whichever other concepts track the reducing entities. But then C is not a basic concept. Q.E.D.

This rationale is bad, however. Grant for the sake of argument that if an entity reduces to other entities, then an essence-tracking concept of the reduced entity must track it *by* being sensitive, somehow, to the presence of the reducing entities.[37] It is still a mistake to think that this sensitivity to the reducing entities must occur *conceptually* rather than *non-conceptually*. To make the point concrete, suppose that the entity at issue is water, which reduces (say) to H_2O. Now imagine a creature that just has brute non-conceptual mechanisms to detect hydrogen, oxygen, and chemical bonds. Whenever these mechanisms identify some substance as H_2O, the creature applies its concept of water to that substance. This creature's concept of water appears to be essence-tracking and yet basic: *essence-tracking*, in virtue of the concept's connection to non-conceptual mechanisms that detect the presence of hydrogen, oxygen, and chemical bonds; *basic*, in virtue of the concept's reference being determined by its connection to these *non-conceptual* mechanisms and not by its connection to any other concepts.

Once we realize that sensitivity to essences can be non-conceptual, we should recognize that the independent plausibility of the semantics-metaphysics link is low.

There is another test worth applying to our two hypotheses: the test of parsimony. The reductivist-friendly hypothesis receives an excellent score on this test. After all, it simply recycles materials that have already been furnished by our best psychological theories. There is much that can be explained via the posit that our psychological capacities are the products of evolution, and more still that can

---

[37] For the record, I reject this principle.

be explained via the posit that our psychologies include a centralized conceptual system as well as many specialized non-conceptual systems. The reductivist-friendly hypothesis simply gives new work to these old posits. By contrast, the anti-reductivist hypothesis is notably profligate. The anti-reductivist hypothesis commits us to an abundance of fresh primitives – phenomenal primitives, normative primitives, and so on. That makes for a poor score on the parsimony test.

In sum, I find the anti-reductivist hypothesis lacking in independent plausibility and lacking in parsimony. These lacks are great; they are not to be outweighed by a small gain in elegance. Between our two contenders, then, I judge the reductivist-friendly hypothesis to be the clear winner.

5.3. Idealizing the semantics-metaphysics link

At this point, the anti-reductivist may want to try a more radical change in tactics. Thus far, we have been evaluating the

> **Semantics-metaphysics link**: Any essence-tracking basic concept must refer to a metaphysical primitive.

This principle posits a link between facts about metaphysics and facts about our *actual* conceptual schemes. But perhaps the anti-reductivist will think it better to link facts about metaphysics to facts about *ideal* conceptual schemes, like so:

> **Idealized semantics-metaphysics link**: Within any ideal conceptual scheme, any essence-tracking basic concept must refer to a metaphysical primitive.[38]

The idea would be that meeting this condition is part of what it *is* for a conceptual scheme to be ideal, in the relevant sense. If we like, we may add further stipulations about the ideal in question: we may stipulate that within any ideal conceptual scheme, there must be a one-to-one mapping between metaphysical primitives and essence-tracking basic concepts; similarly, we may stipulate that within any ideal conceptual scheme, there must be a one-to-one mapping between (reasonably natural?) metaphysical *non*-primitives and essence-tracking *non*-basic concepts.

The move to ideal conceptual schemes solves the problem of independent plausibility: it *is* independently plausible that within any ideal conceptual scheme, any essence-tracking basic concept must refer to a metaphysical primitive. For a conceptual scheme that fails to meet this condition thereby fails to *count* as ideal, in the present sense of that term. The move to ideal conceptual schemes also lets the anti-reductivist agree with me that *for creatures like us*, facts about which essence-tracking concepts are basic are to be explained in psychological and evolutionary terms. But these considerations are not obviously relevant to understanding which essence-tracking concepts would be basic *within any ideal conceptual scheme*. The move to ideal conceptual schemes even lets the anti-

---

[38] [Name removed] made roughly this suggestion to me in conversation.

reductivist allow for the possibility of *gunk* that is phenomenal, normative, intentional, etc. For if there is (e.g.) phenomenal gunk, then arguably an ideal conceptual scheme would have to contain infinitely many phenomenal concepts, none of which would be essence-tracking and basic.[39]

Unfortunately, however, recasting the dispute in terms of the idealized semantics-metaphysics link should give no comfort to the anti-reductivist.

To appreciate why not, we must first properly locate the dispute between the reductivist and the anti-reductivist vis-à-vis this new linking principle. The dispute should no longer be understood as a dispute over the linking principle itself, for unlike the original semantics-metaphysics link, the idealized semantics-metaphysics link is a tautology. An ideal conceptual scheme *just is* a conceptual scheme in which – perhaps among other things – every essence-tracking basic concept refers to a metaphysical primitive. (The anti-reductivist has made her linking principle plausible only by emptying it of content.)

Thus, in relation to the idealized metaphysics-link, the dispute between the reductivist and the anti-reductivist must be located further down the line: the anti-reductivist should assert that the ideal conceptual scheme must include essence-tracking basic concepts for certain phenomenal entities, normative entities, etc., and the reductivist should deny this.

But – and this is the pivotal question – what dialectically admissible evidence can the anti-reductivist give us for taking her view of what an ideal conceptual scheme would look like? I say that she has none.

She cannot examine any pre-existing ideal conceptual schemes – we have none lying around. She cannot just point to facts about what is metaphysically primitive – those facts are precisely the ones at issue, so such assertions would shamelessly beg the question. And she cannot appeal to the fact that the explanatory gap phenomenon arises in certain domains *for beings like us* – we have already seen that this fact, which pertains to our *actual* conceptual schemes, is best explained by a reductivist-friendly hypothesis.

So the idealized semantics-metaphysics link is true, but its truth does not help the anti-reductivist. The semantics-metaphysics link *would* help the anti-reductivist, if only it were true; but according to our best explanatory hypothesis, it is not.

6. An unsound objection

So far I have met the anti-reductivist's general explanatory gap argument with an equally general semantic diagnosis. Yet my diagnosis may appear to *over*generalize. We must be able to disconfirm false reductive claims and confirm true ones, but how are we to do so without appeal to explanatory gaps?

---

[39] Incidentally, this might seem to pose a problem for my diagnosis of the explanatory gap in terms of essence-tracking basic concepts: arguably a phenomenal explanatory gap might yet arise within an ideal conceptual scheme like this, which contains infinitely many phenomenal concepts none of which are essence-tracking and basic. But there is no problem here: my diagnosis of the explanatory gap phenomenon is meant to apply only to creatures – like us – who have finitely many concepts.

The question is pressing only if it is about reductive claims concerning *gap-generating* entities. After all, I have conceded for the sake of argument that all other reductions are transparent, just as the anti-reductivist claims. Still, the question stands: *in these special cases*, how is confirmation or disconfirmation to be had?

By perfectly ordinary means, I reply. To take a silly example, suppose I wish to show that the property of being good does not reduce to the property of being a teapot. Many simple arguments would suffice, such as the argument that some actions are good, but no actions are teapots. Or, shifting to a serious example, I might hypothesize that certain experiences (experiences of phenomenal redness, of phenomenal greenness, etc.) reduce to certain neural firing patterns. To support my hypothesis, I might first identify distinctive properties of and relations among these experiences – e.g., that experiences of phenomenal colors help explain some of my discriminatory capacities, that certain sorts of experiences of phenomenal redness and of phenomenal greenness are mutually exclusive, and that experiences of phenomenal redness are more similar to experiences of phenomenal orangeness than to experiences of phenomenal yellowness. I might then argue that the respective neural firing patterns have the same properties and stand in the same relations. In short, when evaluating a reductive claim involving the referent of an essence-tracking basic concept, we need only draw on garden-variety epistemic methods, including deductive argumentation, inference to the best explanation, and probabilistic reasoning.[40]

If anti-reductivism is to find support, it will be on the basis of these ordinary epistemic methods. It will not be on the basis of the explanatory gap phenomenon.

7. Conclusion

The explanatory gap phenomenon is general, I have assumed, and warrants a general diagnosis. The anti-reductivist and I each have one to offer. My diagnosis is *purely semantic*: I say that an explanatory gap is a symptom of an essence-tracking basic concept within the relevant conceptual family, and nothing more. The anti-reductivist's diagnosis is *semantic-cum-metaphysical*: she accepts my diagnosis, but adds that any essence-tracking basic concept must refer to a metaphysical primitive. My central conclusion is that the purely semantic diagnosis is clearly superior to its semantic-cum-metaphysical rival.

My purely semantic diagnosis is general. But it is also flexible. Because it lays down only one key constraint, the constraint that any gap-generating concept must be an essence-tracking basic concept, it is compatible with a wide range of reductivist proposals about what the referents are, and how they are secured, for phenomenal, normative, and other gap-generating concepts. The diagnosis is even compatible with the possibility that different essence-tracking basic concepts secure their referents in different ways: in a slogan, which I express with apologies to Tolstoy, perhaps *all non-basic essence-tracking concepts are alike, but each basic essence-tracking concept is basic in its own way*. When considering the phenomenal

---

[40] See Mehta (2013b) and Mehta (2014), for example.

concept strategy, the expressivist strategy, and other strategies of the same ilk, we may therefore mix and match as we please to account for the peculiarities of each case. Yet this is not *ad hoc*: we do not appeal to the idiosyncrasies of these concepts, but only to a feature that they share, to explain why we find all of the explanatory gaps that we do.

By offering the purely semantic diagnosis, I have thus not discarded the fragmented reductivist-friendly diagnoses of previous thinkers. Far from it! – I have attempted to forge these pieces into a seamless whole.

REFERENCES

Altman, A. (2004). "Breathing life into a dead argument: G.E. Moore and the open question." *Philosophical Studies* 117 (3): 395-408.

Bedke, M. (2012). "Against normative naturalism." *Australasian Journal of Philosophy* 90 (1): 111-129.

Blackburn, S. (1993). *Essays in Quasi-Realism.* New York: Oxford University Press.

Blackburn, S. (1998). *Ruling Passions.* Oxford: Clarendon Press.

Bliss, R. (2014). "Circles of ground." *Metaphilosophy* 45 (2): 245-256.

Block, N. (1996). "Mental paint and mental latex." *Philosophical Issues* 7: 19-49.

Block, N. and Stalnaker, R. (1999). "Conceptual analysis, dualism, and the explanatory gap." *Philosophical Review* 108: 1-46. Cornell University Press. Reprinted in *Philosophy of Mind: Classical and Contemporary Readings*, ed. D. Chalmers (New York: Oxford University Press, 2002), 371-394.

Braddon-Mitchell, D. (2003). "Qualia and analytical conditionals." *Journal of Philosophy* 100 (3): 111-135.

Burge, T. (1979). "Individualism and the mental." *Midwest Studies in Philosophy* 4: 73-121.

Byrne, A. (1999). "Cosmic hermeneutics." *Philosophical Perspectives* 13: 347-383.

Campbell, J. (2002). *Reference and Consciousness.* Oxford: Oxford University Press.

Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory.* New York: Cambridge University Press.

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

Chalmers, D. (2003). "The content and epistemology of phenomenal belief." In *Consciousness: New Philosophical Perspectives*, eds. Q. Smith and A. Jokic (Oxford University Press: 2003).

Chalmers, D. (2004). "Imagination, indexicality, intensions." *Philosophy and Phenomenological Research* 68: 182-190.

Chalmers, D. (2007). "Phenomenal concepts and the explanatory gap." In *Phenomenal Concepts and Phenomenal Knowledge*, eds. T. Alter and S. Walter (New York: Oxford University Press), pp. 167-194.

Chalmers, D. (2012). *Constructing the World.* New York: Oxford University Press.

Chalmers, D. and Jackson, F. (2001). "Conceptual analysis and reductive explanation." *Philosophical Review* 110: 315-361.

Diaz-Leon, E. (2011). "Reductive explanation, concepts, and a priori entailment." *Philosophical Studies* 155 (1): 99-116.

Dretske, F. (1995). *Naturalizing the Mind.* Cambridge, Massachusetts: MIT Press.

Fine, K. (1994a). "Essence and modality." *Philosophical Perspectives* 8: 1-16.

Fine, K. (1994b). "Senses of essence." In *Modality, Morality and Belief*, eds. W. Sinnott-Armstrong, D. Raffman, and N. Asher (Cambridge: Cambridge University Press).

Fine, K. (2015). "Unified foundations for essence and ground." *Journal of the American Philosophical Association* 1 (2): 296-311.

Fish, W. (2008). "Relationalism and the problems of consciousness." *Teorema* 28 (3): 167-80.

Fish, W. (2009). *Perception, Hallucination, and Illusion.* Oxford: Oxford University Press.

Fodor, J. (1990). *A Theory of Content and Other Essays.* Cambridge, Massachusetts: MIT Press.

Gibbard, A. (1990). *Wise Choices, Apt Feelings.* Cambridge, Massachusetts: Harvard University Press.

Gibbard, A. (2003). *Thinking How to Live.* Cambridge, Massachusetts: Harvard University Press.

Hawthorne, J. (2002). "Advice for Physicalists." *Philosophical Studies* 108: 17-52.

Hill, C. and McLaughlin, B. (1999). "There are fewer things in reality than are dreamt of in Chalmers' philosophy." *Philosophy and Phenomenological Research* 59: 445-454.

Horgan, T. and Timmons, M. (1991). "New wave moral realism meets moral twin earth." *Journal of Philosophical Research* 16: 447-465.

Jackson, F. (1982). "Epiphenomenal qualia." *Philosophical Quarterly* 32: 127-36.

Jackson, F. and Pettit, P. (1995). "Moral functionalism and moral motivation." *Philosophical Quarterly* 45 (178): 20-40.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis.* Oxford: Clarendon Press.

Jackson, F. (2007). "A priori physicalism." In B. McLaughlin and J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind.* Blackwell University Press.

Kalderon, M. (2004). "Open questions and the manifest image." *Philosophy and Phenomenological Research* 68 (2): 251-289.

Leslie, S. (2013). "Essence and natural kinds: When science meets preschooler intuition." *Oxford Studies in Epistemology* 4: 108-66.

Levine, J. (1983). "Materialism and qualia: the explanatory gap." *Pacific Philosophical Quarterly* 64: 354-361.

Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness.* New York: Oxford University Press.

Levine, J. (2010). "The Q factor: modal rationalism vs. modal autonomism." *Philosophical Review* 119 (3): 365-380.

Levine, J. (2014). "Modality, semantics, and consciousness." *Philosophical Studies* 167: 775-784.

Lewis, D. (1966). "An argument for the identity theory." *Journal of Philosophy* 63: 17-25.

Lewis, D. (1972). "Psychophysical and theoretical identifications." *Australasian Journal of Philosophy* 50 (3): 249-258.

Lewis, D. (1979). "Attitudes *de dicto* and *de se.*" *Philosophical Review* 88: 513-543.

Lewis, D. (1988). "What experience teaches." *Proceedings of the Russellian Society.* Reprinted in *Philosophy of Mind: Classical and Contemporary Readings*, ed. D. Chalmers (New York: Oxford University Press, 2002), 226-246.

Loar, B. (1990/1997). "Phenomenal states." *Philosophical Perspectives* 4: 81-108. Revised version in *The Nature of Consciousness*, eds. N. Block, O. Flanagan, and G. Güzeldere: 597-616. Cambridge: MIT Press, 1997.

Lormand, E. (ms). "How to Be a Meaning Atomist." At http://www-personal.umich.edu/~lormand/phil/meaning/atomism.htm.

Lormand, E. (2004). "The explanatory stopgap." *Philosophical Review* 113: 303-357.

Lycan, W. (1996). *Consciousness and Experience*. Cambridge, Massachusetts: MIT Press.

Martin, M. (2002). "The transparency of experience." *Mind and Language* 17 (4): 376-425.

Martin, M. (2006). "On being alienated." In *Perceptual Experience*, eds. T. Gendler and J. Hawthorne (Oxford University Press), pp. 354-410.

McGinn, C. (1991). *The Problem of Consciousness: Essays toward a* Resolution. Oxford: Basil Blackwell.

Mehta, N. (2013a). "How to explain the explanatory gap." *Dialectica* 67 (2): 117-135.

Mehta, N. (2013b). "Beyond transparency: the spatial argument for experiential externalism." *Philosophers' Imprint* 13 (8): 1-19.

Mehta, N. (2014). "The limited role of particulars in phenomenal experience." *Journal of Philosophy* 111 (6): 311-331.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*. MIT Press.

Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

Nuccetelli, S. and Seay, G. (2007). "What's right with the open question argument." In *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, eds. S. Nuccetelli and G. Seay (Oxford: Oxford University Press).

Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.

Papineau, D. (2011). "What exactly is the explanatory gap?" *Philosophia* 39 (1): 5-19.

Perry, J. (1979). "The problem of the essential indexical." *Noûs* 13: 3-21.

Perry, J. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press.

Pigden, C. (2012). "Identifying goodness." *Australasian Journal of Philosophy* 90 (1): 93-109.

Rosati, C. (1995). "Naturalism, normativity, and the open question argument." *Noûs* 29 (1): 46-70.

Rosen, G. (2010). "Metaphysical dependence: grounding and reduction." In *Modality: Metaphysics, Logic, and Epistemology*, eds. B. Hale and A. Hoffman (Oxford: Oxford University Press).

Schaffer, J. (forthcoming). "The ground between the gaps." *Philosophers' Imprint*.

Schroeder, M. (2005). "Realism and reduction: the quest for robustness." *Philosophers' Imprint* 5 (1): 1-18.

Schroeder, M. (2007). *Slaves of the Passions*. Oxford: Oxford University Press.

Searle, J. (1980). "Minds, brains, and programs." *Behavioral and Brain Sciences* 3: 417-424.

Smith, M. (1994). *The Moral Problem*. Malden, Massachusetts: Blackwell University Press.

Stampe, D. (1977). "Towards a causal theory of linguistic representation." *Midwest Studies in Philosophy* 2 (1): 42-63.

Strandberg, C. (2004). "In defence of the open question argument." *Journal of Ethics* 8 (2): 179-196.

Sturgeon, S. (2008). "Disjunctivism about visual experience." In *Disjunctivism: Perception, Action, Knowledge*, eds. A. Haddock and F. Macpherson (New York: Oxford University Press), pp. 112-143.

Sundström, P. (2017). "How physicalists can – and cannot – explain the seeming "absurdity" of physicalism." *Philosophy and Phenomenological Research* (online first).

Sutton, P. (2014). "Moore's 'new' open question argument." *Res Philosophica* 91 (4): 681-693.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind.* Cambridge, Massachusetts: MIT Press.

Tye, M. (2000). *Consciousness, Color, and Content.* Cambridge, Massachusetts: MIT Press.

Yetter-Chappell, H. and Chappell, R. (2013). "Mind-body meets metaethics: a moral concept strategy." *Philosophical Studies* 165 (3): 865-878.