Ground Functionalism

Jonathan Schaffer, Rutgers University
Forthcoming, *Oxford Studies in Philosophy of Mind*

> O Chestnut tree, great rooted blossomer,
> Are you the leaf, the blossom or the bole?
> O body swayed to music, O brightening glance,
> How can we know the dancer from the dance?
> (William Butler Yeats, *Among School Children*)

Back when the world was young—in the 1960s and 1970s—a hopeful story was told about the metaphysics of the mind, which went something like this. For hundreds of years, we philosophers were held spellbound by Descartes's spooky dualism. Ryle and the behaviorists then freed us from dualism but omitted inner mental life. Place, Smart, and the identity theorists then found room for inner mental life but missed multiple realizability. And finally Putnam, Armstrong, Lewis, Fodor, and the functionalists forged a grand synthesis that reconciled materialism, inner mental life, and multiple realizability, and all was well with the world (or at least with the place of the mind within it). We could rest content with materialism, and move on to consider how connectionism would explain cognition and information theory would naturalize content.

Alas this hopeful story has aged poorly (in many ways). For objections soon arose against functionalism, including Block's objection from overgeneration and Kim's objection from causal potency. But perhaps the most damaging objection, from Kripke, Levine, and Chalmers, was that functionalism—and indeed all materialist-friendly views—could not explain conscious experience but rather fell into an explanatory gap. This objection remains "the main obstacle to acceptance of materialism" (Levine 2001: 76; see also Gertler 2001: 689), making many contemporary metaphysicians of mind return to dualism, or embrace options such as Leibnizian panpsychism or Russellian monism, all of which take phenomenal (or proto-phenomenal) properties to be fundamental ingredients in nature. According to the new tragic story about the metaphysics of the mind, from dualism we came and to dualism we shall return.

I have the bold ambition of reviving the hopeful materialist story, by adding a new chapter—*ground functionalism*—which integrates functionalist insights about the mind with ground-theoretic insights about explanation. The ground functionalist posits a *mind making* principle linking material states to mental states via functional role, such that a properly choreographed system dances out a mind. I argue that ground functionalism preserves the insights of functionalism, while enabling a viable explanation for consciousness.
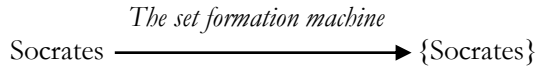
1. Metaphysical Explanation

Lawyers say that hard cases make bad law. Since I am after a viable explanation for the hard case of consciousness, I propose to first consider how metaphysical explanation works in some easier cases, in order to trace an explanatory template (applied to consciousness in §2). On this template, there is a viable form of metaphysical explanation linked by grounding principles, aptly modeled by structural equations, and justifiably posited when explanatorily fruitful.

1.1 Four examples

In order to trace a template for metaphysical explanation, I offer four starting point examples, drawn from set theory, mereology, metaethics, and quantum mechanics. Each example involves controversial assumptions, but the assumptions are not at issue. Rather what is at issue are the explanations they would enable. So for instance, even if you do not believe in sets, you can still consider what explanations set theory would enable. Call an explanation *viable* when it is successful given its background assumptions. I claim that each example displays a viable metaphysical explanation *linked by a grounding principle*.

So starting with the example drawn from set theory, consider the relationship between Socrates and {Socrates}, the set whose one and only member is Socrates. I assume that {Socrates} is built out of Socrates by the principle of set formation (or set making). We can think of set formation as a machine that clamps things into sets. For every plurality of inputs one feeds into this machine, one gets out a set with all and only those inputs as members. So if one feeds Socrates into this machine, one gets out {Socrates}:[1]

The set formation machine
Socrates ──────────────────▶ {Socrates}

My first claim about this example is that there is a viable metaphysical explanation for why {Socrates} exists, from the existence of Socrates, via the principle of set formation. After all, given that Socrates exists and set formation is at work, it is no mere accident that {Socrates} exists too. For there is a grounding principle: set formation, which clamps every plurality of inputs into a so-membered set (never mind how set formation came to operate—we are just assuming it is at work). And there is an input: Socrates (never mind where Socrates came from—we are just assuming he is there). {Socrates} is the output of this principle on this input, and so {Socrates} exists.

To buttress this first claim, note that the theoretical role of explanation includes (i) *revealing patterns*, (ii) *providing recipes*, and (iii) *allowing understanding* (Schaffer 2018: 305–07). Seeing {Socrates} as the output of Socrates via set formation (i) reveals a unifying pattern that extends through Plato and {Plato}, Aristotle and {Aristotle}, etc.; (ii) conveys a recipe to wiggle the existence of {Socrates} by wiggling the existence of Socrates; and (iii) allows one to understand why {Socrates} exists, by revealing how sets are collected.

My second claim is that set formation plays the needed role of *link* from Socrates to {Socrates}. If one just considered Socrates and {Socrates} on their own, one would just be considering a heap of things with no specified relationship (likewise for the existence facts concerning these things—this is just a heap of facts). That is not enough for explanation. Explanation requires *connection* as well as *direction*. Set formation provides the direct connection from Socrates to {Socrates}, so licensing the arrow in the diagram above.

Turning to a second example drawn from mereology, consider the relationship between some appropriately bonded H, H, and O atoms and the H_2O molecule they compose. I assume that the molecule is built out of the atoms by the principle of mereological fusion (or whole making). We can think of mereological fusion as a machine that melts things into sums. For every plurality of inputs one feeds into this machine, it spits out their sum. So if one feeds the H, H, and O atoms into the machine, it spits out their sum, namely the H_2O molecule:[2,3]

---

[1] Strictly speaking, this is a derivative principle describing the operation of hierarchical set formation *at a stage*. Hierarchical set formation is *recursive*, in that the outputs at any given stage loop back into the mouth of the machine to serve as added inputs at the next stage. So if we start with Socrates at stage 0, then we get {Socrates} born at stage 1, {{Socrates}} born at stage 2, etc. Note that, although I am only concerned with a single stage of set formation in the main text, the total recursive structure of the machinery further justifies the claim of *priority*, that Socrates comes first and {Socrates} second (with {{Socrates}} third, etc.).

[2] This is the universal fusion machine. For restricted fusion machines, one needs restrictions on the inputs. For instance, for the organicist fusion machine (van Inwagen 1990), one needs to say: "For every plurality of inputs *whose activities collectively constitute a life* that one feeds into this machine…" But such a machine will not spit out any H_2O molecules. For the example, any fusion machine that can make H_2O molecules will do.

[3] *Clarification*: The fusion machine spits out a sum, but falls silent as to whether that sum is a cabbage or a king, etc. What makes the sum *an H_2O molecule* presumably involves extra-mereological facts as to the natures of and relations between the parts (for instance that the sum has an exhaustive decomposition into two H atoms and one O atom, where the O atom is covalently bonded to each H atom). That is not at issue here.

The mereological fusion machine

H
H ———————————————→ H2O
O
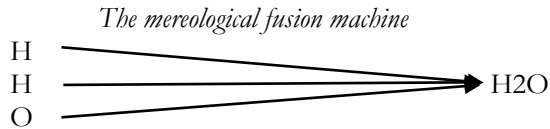
There is a viable metaphysical explanation for why the sum (the H2O molecule) exists, from the existence of the parts (the H, H, and O atoms), via mereological fusion. After all, given that the parts exist and the principle of mereological fusion is at work, it is no coincidence that the sum exists too. The sum is the output of this principle on this input. Mereological fusion provides a unifying pattern which conveys recipes and provides a basis for understanding, revealing how sums are lumped.

Mereological fusion plays the needed role of *link* from atoms to molecule. Without it one would just see a list of things (/facts). Mereological fusion contributes the directed connection needed for explanation, underwriting the arrows.

Moving to a third example drawn from meta-ethics, consider a simple "do as you like" subjectivist treatment of moral rightness, on which the present desires of the actor determine the moral status of her act. We can think of this subjectivist as positing an operation of right making, understood as a machine which gilds acts with morality. Feed it the present desires of the actor, and its spits out the moral status of her act:[4]

The right making machine

Desires ———————————————→ Moral status
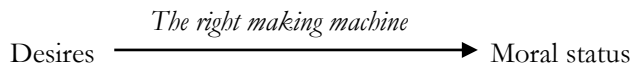
There is a viable metaphysical explanation for why Ann's eating the chocolate is right, from Ann's desire to eat the chocolate, via right making. After all, given that Ann desires to eat the chocolate and the principle of right making is at work, it is hardly surprising that Ann's eating the chocolate is right. The moral status of the act is the output of this principle on this input. Right making provides a unifying pattern, conveys recipes, and allows understanding, revealing how moral status is painted.

Right making plays the needed role of *link* from desires to moral status. Without it one would just see a non-explanatory list of states (/facts). With it we see a directed connection.

Finishing with a fourth example drawn from quantum mechanics, consider the treatment of particle locations afforded by Bohmian wave function realism (Albert 1996), on which what is fundamental is a wave in a $3n$-dimensional space and a single world-atom (the "marvelous point") in $3n$-space, and what is derivative are n-many "ordinary" physical particles in the 3-space we experience, as derived by the operation of particle enactment (or particle placing). We can think of particle enactment as a machine that projects particles onto 3-space. Feed it the location of the world in $3n$-space at a given time, and it spits out 3-space locations for n particles at that time, by assigning the first triple of the world's $3n$-space location as the first particle's 3-space location, the second triple of the world's $3n$-space location as the second particle's 3-space location, etc. So if one feeds the location of the world in the $3n$-space at a given time into the machine, it spits out locations for

---

[4] I trust the reader can see how to extend this simple subjectivism to more sophisticated views. For instance, on Railton's (1986: 173–74) ideal advisor approach, what makes an action right for an agent is not her present desires, but what an idealized "A+" version of himself "would want his non-idealized self A to want."

n-many particles in 3-space at that time, with particle p_j placed at coordinates $\langle x, y, z\rangle$ in 3-space, corresponding to the world's $\langle 3j\text{-}2, 3j\text{-}1, 3j\rangle$ coordinates in $3n$-space:[5,6]

The particle enactment machine
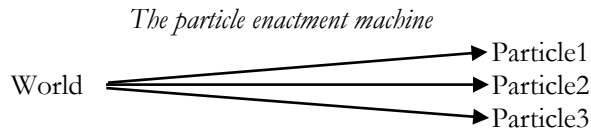
There is a viable metaphysical explanation for where the particles are, from the location of the world in 3-space, via particle enactment. After all, given that the world is there and the principle of particle enactment is at work, it is no wonder that the particles are where they are. The locations of the n particles in 3-space at that time are the output of this principle on this input. Particle enactment provides a unifying pattern, conveys recipes, and allows understanding, revealing how particles are sprayed.

Particle enactment plays the needed role of *link*, arrowing world-location to particle-locations.

Bringing this together, I have offered four examples of viable metaphysical explanations, each linked by a grounding principle. If I am right about any of these cases, then:

Minimal explanatory claim: There are viable metaphysical explanations linked by grounding principles

Note that I am helping myself to some fairly orthodox ideology, including that of "metaphysical explanation," and "grounding principle," but this ideology is not load-bearing. Mainly I am just looking at some non-causal but constitutive explanations and labeling their linking principles (in order to set up an analogous explanation for consciousness). The reader is invited to substitute terminology as she prefers.[7]

1.2 Theorizing the examples
So far I have displayed four viable metaphysical explanations linked by grounding principles. Now I offer an explanatory template based on these four examples and an example of causal explanation. This template has a tripartite structure, where *sources* (causes, grounds) via *link* (causal law, grounding principle) generate *result* (effect, derivative), aptly modeled by the input-function-output structure of *structural equation models*.

[5] Albert (1996; see also Ney 2013) calls this a principle of enactment because, under Bohmian wave function realism, having the world particle at such coordinates enacts the causal roles of having so-many ordinary particles so-located. Note that this implementation does not eliminate particles in three-dimensional space but embraces them as derivatively real, and it does not treat particle enactment as a follow-on that must be justified from the preceding theory but rather as one of the starting point posits of the theory.
[6] Why do I bother with such a complicated example? Because it has many useful features, including (i) it is scientifically serious, (ii) it is one-many, (iii) it is broadly functionalist, (iv) it is a case where some commentators (such as Maudlin 2010 and Allori 2013) have questioned how an explanation is possible, and (v) it is a case in which the need for a non-transparent grounding principle is especially vivid (§1.3).
[7] My terminology is drawn from Fine (2001), Schaffer (2009), Rosen (2010), and the grounding theorists, and most especially follows Schaffer (2016, 2017, 2018). But I make no specific assumptions about grounding. Indeed even *grounding skeptics* such as Wilson (2014) can accept the substance of what I am saying, but should just substitute the more specific metaphysical determination relation of 'enactment' where I use the generic notion of 'grounding'. Also I make no further assumptions about linking principles, beyond how they provide explanatory connections. There is a division among grounding theorists—discussed by Glazier (2020: §3)—between *nomicists* (such as myself and Wilsch 2015) who invoke lawlike principles, and *essentialists* (such as Fine 2012 and Dasgupta 2014) who invoke the essences of the derivative. Essentialists should just substitute claims about the essence of the derivative where I invoke a principle. (I say how to re-write ground functionalism in both Wilsonian and Finean terms in §2.1.)

So consider causal explanation. One traces how causes generate an effect via laws of nature. To take a paradigmatic example, imagine that Suki throws a rock through a window. We can think of the shattering of the window as the result of Suki's throwing the rock, via the nomological machinery.[8] If one feeds Suki's throw into the causal machinery of the world, one gets out a shattered window:

$$\text{Throw} \xrightarrow{\textit{The window shattering law}} \text{Shattering}$$

I say that—on this picture—there is an explanation for why the window shattered, from Suki's throw, via the "window shattering law." After all, given that Suki threw the rock and the window shattering law is at work, it is no coincidence that the window shattered. Rather the shattering is the output of this law on this input. This is not just a paradigmatic example, but one that unifies a pattern across vandalisms, conveys recipes for wiggling the fates of windows, and offers a basis for understanding why the window shattered.

With causal explanation, laws of nature provide the needed links. Without them one would just see a list of events. The causal laws add the connection and direction that constitute the arrow of explanation.

Indeed in the causal case, sometimes cause and effect are known but the law is not—and then the effect remains unexplained. Consider the classic double slit experiment (from Thomas Young in 1801), in which light shone through a doubly-slit wall generates an interference pattern on a screen, which Feynman (1965: 1.1) calls "impossible, *absolutely* impossible, to explain in any classical way," saying that it "has in it the heart of quantum mechanics." Here cause (shining the light through the doubly-slit wall) and effect (interference pattern on the screen) were long known, but the explanation required the discovery of quantum mechanical laws providing the link.

So putting the examples together, here is the template I have in mind:

Explanatory template: If the result is the output of the sources via the link, then there is a viable explanation of the result from the sources via the link

Note that *Explanatory template* is merely a sufficient condition for a viable explanation, applicable only in simple cases where the connections are direct. There are unresolved difficulties arising in more complex cases with various indirect connections, but these need not not intrude here.

The reader familiar with *structural equation models* (Pearl 2000, Woodward 2003) will recognize my talk of "input," "function," and "output," and the directed graphs I draw. These models were developed for studying causation but there is nothing specifically causal about them. The formalism smoothly extends to grounding (Schaffer 2016). In a nutshell, we recruit variables whose valuables represent the states of interest for the system under study, link the dependent variables by functions that map the values of some input variables to their output value, and then assign values to the independent variables.[9]

So for a simple model of Suki throwing the rock through the window, we can recruit an independent variable *Throw* with values of 0 if Suki does not throw and 1 if she throws, and a dependent variable *Shatter* with values of 0 if the window does not shatter and 1 if the window shatters. We then link these variables

---

[8] To treat the example more seriously, we either need to specify the precise background conditions around Suki's throw and the precise details of her throw, or—more feasibly, as will be done in the main text—we need to use a locally valid but derivative generalization linking the rock throw to the window shattering.

[9] This is how it goes in the deterministic case. In the indeterministic case, the functions output probability distributions over values for each dependent variable, and we assign values to all variables consistent with what the functions allow. Since I take it that the metaphysical case must be deterministic, I focus there.

with a function—here the identity function—mapping *Throw*=0 to *Shatter*=0 and *Throw*=1 to *Shatter*=1. Finally we assign *Throw*=1, and now we can solve for the fate of the window, since we can derive *Shatter*=1.

More formally, we begin with a representation of the system under study, divided into sets of independent and dependent variables, with each variable mapped to some range of allotted values:

*Variables1*: <*Independents1*={*Throw*}, *Dependents1*={*Shatter*}, *Range1*={*Throw*→{0,1}, *Shatter*→{0,1}}>

We then add in functions given the "dynamics" of the system, where '<=' is to be read as 'is the output of' (yielding '*Shatter* is the output of the identity function on *Throw*'):

Functions1: {*Shatter* <= *Identity*(*Throw*)}

Finally—in the deterministic case—we just need to seed the "initial conditions" by assigning values to the independent variables:

Assignment1: {*Throw*=1}

This structure induces the following directed graph, fitting our earlier representation:

Throw ⎯⎯⎯⎯⎯⎯⎯⎯▶ *Shatter*
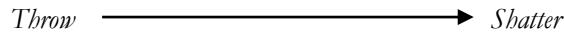
But we have added needed information beyond the graph, concerning the range of values each variable can take, the form of the function connecting them, and the value assigned to the independent variables. This added information enables us to derive *Shatter*=1 (as the output of the identity function on the input *Throw*=1), and so model a viable explanation as to why the window shattered.

For a metaphysical example, we can simply re-purpose the model for the window shattering, to represent the formation of {Socrates}. Just treat *Throw* as a strangely labeled variable for Socrates, with 0 representing non-existence and 1 existence; and treat *Shatter* as a strangely labeled variable for {Socrates}, with 0 again representing non-existence and 1 existence. We can then solve for the presence of {Socrates}, and in the exact same way display an explanatory model as to why {Socrates} exists. So we find a common modeling structure for causal and metaphysical explanations, from sources (Suki's throw, Socrates's existence), via link (window shattering law, set formation principle), to result (the shattering, {Socrates}'s existence).

Or to add a new metaphysical example, suppose we want to model how the mass of the H2O molecule depends on the mass of its constituent H, H, and O atoms. Then we could recruit variables *H1*, *H2*, *O*, and *H2O* (representing the atoms and molecule), allowed any real value (representing mass in Daltons). We would then need to specify the function through which the mass of the constituent atoms determines the mass of the molecule. Supposing a Newtonian framework, the apt function is addition. And finally we would need to assign masses in Daltons to the atoms, as initial inputs. Rounding off a bit, we can assign *H1* and *H2* the value of 1, and *O* the value of 16. We would then be able to solve for the mass of the H2O molecule, since we can now derive *H2O*=18. The model would then look like this:

Variables2: <*Independents2*={*H1*, *H2*, *O*}, *Dependents2*={*H2O*}, *Range2*={*H1*/*H2*/*O*/*H2O*→*Reals*}>
Functions2: {*H2O* <= *Addition*(*H1*, *H2*, *O*)}
Assignment2: {*H1*=1, *H2*=1, *O*=16}

The model induces the same directed graph as the mereological fusion function displayed in §1.1:

$$H1 \quad\rightarrow$$
$$H2 \quad\longrightarrow\quad H2O$$
$$O \quad\nearrow$$
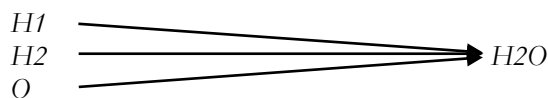
But the model adds further information beyond the graph, such as the substantive information that mass is additive (which is true in a Newtonian framework but false in relativistic physics), which thereby entails $H2O$=18, aptly representing the mass of the molecule as (roughly) 18 Daltons. In this way we can model a viable Newtonian explanation for why the mass of the H2O molecule is (roughly) 18 Daltons.

The structural equation formalism helps reveal exactly why explanations need links. The links not only license the arrows, but they license specific structural equations that enable solutions for the values of dependent variables. Take the mass case. We not only need to draw an explanatory arrow from $H1$, $H2$, and O to $H2O$, but we need to specify that it is the arrow of *addition* to solve for the mass of the molecule (had we multiplied, or taken the mean or the min, etc. we would have reached a different result).

There is a familiar saying in computing: "Garbage in, garbage out." I offer a further motto: "No function, no output." From just the input (causes, grounds) nothing further follows.[10]

So our explanatory template can be modeled as follows:

Template modeled: If there is an apt structural equation model in which the value of the variable representing the result is the output of the values of the variables representing the sources on a given function representing the link, then there is a viable explanation of result from sources and link

Note that *Template modeled* (like *Explanatory template*) is merely a sufficient condition for a viable explanation, applicable only in simple cases with a single direct link. Also note that I help myself to the ideology of an "apt" model. There are many vexed questions as to what exactly aptness involves, but I trust that the models in use should count as apt however that goes. The reader who suspects otherwise is invited to say why.

1.3 Abductive metaphysics
The topic of metaphysical explanation is not only relevant to explanatory concerns about consciousness, but goes to the heart of metaphysics itself. I say that a satisfying metaphysics should be *explanatory*. A metaphysics that merely lists what there is (as per Quine 1948) lacks explanatory value, since a mere list is not an explanatory structure and provides no understanding (Jackson 1998: 4–5). Rather, the metaphysician with the ambition to explain *how and why the world is as it is* should offer theories of what grounds what (Schaffer 2009).

I can now be more precise about what it takes to offer a theory of what grounds what. First, one needs to think of the derivative result as enfolded in a contrast space, or (equivalently) encased in a question. This is the analogue of encoding a range for a variable. When it comes to the H2O molecule for example, we might enfold it in a contrast space of {exists, does not exist}, which is to encase it in the question of whether it exists or not. Or we might enfold it in the contrast space of {has mass *n* Daltons} for all real numbers *n*,

---

[10] Likewise from just a function, nothing further follows. There is only output when input meets function. Note that this is compatible with the idea that the input in our model may represent a null condition in the world, and so compatible with Fine's (2012: 47–8) idea of *zero-grounding*, understood as the output of an operation when fed nothing, and illustrated by the idea that the null set is the output of set formation when fed nothing. To model zero-grounding (if such be wanted), we just require an input value that aptly represents the null condition. Actually we can re-purpose our simple two-binary-variable model used to represent Suki throwing the rock through the window, and Socrates membering {Socrates}, by taking the independent variable to represent the null condition, and the dependent variable to represent the null set. (This requires the representation of counterlogical conditions, such as the absence of even the null condition. But structural equation models of grounding were already knee-deep in counterlogicals: Schaffer 2016: 71–3.)

which is to encase it in the question of how massive is it. Secondly, we need to identify the candidate explaining entities, and enfold these in a contrast space/question as well. So we might look to the H, H, and O atoms, and ask of each whether it exists or not, or how massive is it, etc.

Once we have asked our questions of nature, we may consider how the answer to a given question is determined by answers to upstream questions, and ultimately by answers to certain fundamental (/source) questions. This is to exhibit how wiggling the grounds (within its contrast space) wiggles the grounded (within its contrast space), and how the grounded fits into a pattern of dependence upon its grounds. And then we only need to answer the source questions, and we will thereby be in a position not merely *to answer* all downstream questions, but *to understand why these are the answers*, by seeing how answers cascade through nature.

Explanations are not merely answers to a question, but flowcharts as to how answers cascade through nature.

Think of the physicist, tasked with creating an explanatory causal map of the cosmos. She may begin from a statespace representing the possible states of the system at a time. This is a contrast space, posing the question of *how the system is* for any given time. She may then add a dynamics representing the possible trajectories through this statepace over time. This is a function from the state of the system at a time to the state of the system over other times.11 Then she only needs to answer *what was the initial state*. This will position her (in principle) to give a complete causal explanation for the rest of the cosmos, since the dynamics not only determines the subsequent history, but charts how answers evolve forwards through time.

Think of the metaphysician as facing an analogous explanatory task. She needs to answer the question *what is the fundamental state* and specify the "dynamics" of grounding, to chart how answers percolate upwards through the derivative. (So in the case of consciousness, the materialist is looking to answer the question *what is the material state* and specify the dynamics of grounding, to chart how answers to the material state question percolate upwards to determine answers to the question *what is the mental state*.)

I can also be somewhat more precise about the methodology of such an explanatory metaphysics, which is *abductive*. Grounding principles are posited by inference to the best explanation, in a holistic and fallible manner. The idea that principles are posited for explanation is arguably already present in Aristotle's *Posterior Analytics*, where—at least according to Kosman (1973: 387)—Aristotle holds that the "criteria of adequacy" for principles is our ability "to explain by them the phenomena with which we begin and thus to gain with them scientific understanding…" And the idea that metaphysics is generally abductive remains prominent, for instance in Lewis & Lewis's (1970: 212) view of metaphysical debates as "haggling" over the costs and benefits of consistent options, and in Paul's (2012: 23) approach on which "modeling and inference to the best explanation… form the core of the metaphysical method."

I am saying that, just as we accept an abductive methodology for inquiry into the laws of nature, so we should accept an abductive methodology for inquiry into the grounding principles (or "laws of metaphysics"). In the causal case, the physicist seeks the simplest (/most elegant) total package of initial conditions and dynamics, that is the strongest (/most informative) for the rest. Likewise I say that the metaphysician should seek the simplest and strongest package of fundamental conditions and grounding principles. From an evidential perspective, in both physics and metaphysics we are most directly witness to a small swath of intermediate structure—some local patterns in the midst of history at middling levels—and (crudely speaking) our task is to construct the best explanation of this "data," from before and below.12

---

11 Or in the indeterministic case, this is a function from the state of the system at a time to a probability distribution over other times. At present I am only considering the deterministic case.
12 A less crude account should allow us to "explain away" a seeming bit of "data" as a mere appearance, etc.

Inference to the best explanation is ultimately a holistic comparison between packages, which means that any posited links—causal or metaphysical—are *doubly provisional*. It is provisional that a given posited link is the best complement to a given remainder, and provisional that the resulting package provides the best explanation overall. But a physicist can still provisionally recommend a nomic posit (e.g. Schrödinger's equation) when it seems *explanatorily fruitful* (e.g, explaining the result of the double split experiment, etc.), without specifying every other law of nature or considering every alternative package. Likewise a metaphysician can still provisionally recommend a grounding principle (such as set formation) on the basis that it seems *explanatorily fruitful*, without specifying every other grounding principle or considering every alternative package. (It is in this doubly provisional spirit that I recommend ground functionalism.)

But an abductive methodology stands opposed to the constraint that animates the explanatory gap literature, namely that explanatory links be *epistemically transparent*, in the sense of being evident to pure reason, *a priori* knowable, or inconceivable otherwise. Transparency is neither holistic nor abductive, but local and deductive, requiring of each individual link that it be perfectly evident to pure reason. Indeed the opposition between an abductive and rationalist-deductive methodology may lie at the root of the whole debate. Thus Trogdon (2013: 471) connects the explanatory gap concern to a transparency requirement:

> [T]he appearance that connecting questions [between phenomenal and physical facts] are cognitively significant is at the root of a familiar challenge to physicalism. A standard gloss of physicalism is that in the actual world each mental fact is ultimately grounded in certain physical facts… Many agree that the appearance of cognitive significance with respect to connecting questions in the phenomenal/physical case—what Levine dubs the *explanatory gap*—gives us a *prima facie* reason to believe that physicalism is false, that the physical fails to ground the mental.

(I reply that we should not reject physicalism but instead reject transparency, and thus make room for a physicalism that posits non-transparent (/"cognitively significant") connections on abductive grounds.)

Note that there are many ways to precisify transparency. For instance one might require an *a priori* entailment from the grounds to the grounded, so that an ideal reasoner in her armchair, given the information about the grounds and full conceptual competence, could deduce the facts about the grounded on that basis (Chalmers & Jackson 2001: 351). Or—perhaps equivalently—that it should be *inconceivable* for the link to fail, so that an ideal reasoner would regard any other view as contradictory (Chalmers 1996: 73–76).[13] Chalmers (2012) considers a wide range of "scrutability" theses. Nothing here should turn on the precisifications however, so I will simply use "transparency" along with a conceivability test for definiteness.

I oppose transparency for three reasons, the first of which being that it fits none of our starting examples (§1.1). It is fully conceivable that set formation is not at work to clamp things into sets, and likewise it is fully conceivable that mereological fusion, right making, and particle enactment are not functioning. Indeed each of these principles is rejected by various excellent philosophers, who may reject the result as well (e.g. denying that there are any sets), or say that the result is grounded in other ways or else deem it fundamental (e.g. saying that moral status is grounded in non-psychological ways, or viewing it as

---

[13] Chalmers (1996: 73) gives the following as his leading example of such an "inconceivability":
> What kind of world could be identical to ours in every last microphysical fact but be biologically distinct? Say a wombat has two children in our world. The physical facts about our world will include facts about the distribution of every particle in the spatiotemporal hunk corresponding to the wombat, and its children, and their environments, and their evolutionary histories. If a world shared those physical facts with ours, but was not a world in which the wombat had two children, what could that difference consist in? Such a world seems quite inconceivable.

I reply (Schaffer 2017; see also Montero 2013: 102) that one conceivable difference concerns whether or not mereological fusion is at work. There is a conceivable *nihilist* scenario in which the particles do not fuse at all, not even into wombats.

fundamental), or even invert the proposed arrows (e.g. saying that the whole grounds the parts). It is now widely accepted in each of these debates that there are multiple consistent options for the haggling. (Of course it is epistemically possible that, in every single one of these debates, some subtle and hitherto unseen contradiction is lurking in all but one of the relevant options. But this seems unlikely.)14

My second reason for opposing transparency is that it manifestly fails for causal links. In the early days of modern science, there was a felt demand for transparent laws of nature ("rational mechanics"). But we have long since learned to reject any such constraints on laws of nature. So why impose them on grounding principles? The causal case already shows that viable explanation does not need transparency.

My third reason for opposing transparency is that—at least from a broadly realist perspective—causal laws and grounding principles are both principles of real generation in nature (with explanations charting how answers cascade through nature). From such a realist perspective, it seems simply wrong-headed to demand that the clockwork of nature be transparent to pure reason.

Putting this together: Explanatory metaphysics—at least when the explanations fit *Template and Template modeled*—must articulate a structure of sources (/inputs), link (/function), and result (/output). The proper methodology—in both the causal and the metaphysical case—for posting such links is abductive:

> *Abductive methodology*: Links for causal and for metaphysical explanations may be posited when explanatorily fruitful (and ultimately insofar as they are part of the best explanatory package)

Note that rejecting transparency is not saying *anything goes*. The rejection of one (unsuitable) constraint is not the rejection of all constraints. I am saying that, in both physics and metaphysics, the proper constraint on positing links is not epistemic transparency but explanatory fruitfulness.15

2. Ground Functionalism

In §1 I argued that there is a viable form of metaphysical explanation linked by grounding principles, which are aptly modeled as functions from input to output, and justifiably posited when explanatorily fruitful. I now introduce the ground functionalist approach, positing a *mind making* principle mapping token realizer states to token mental states, so enabling a viable materialist-friendly explanation for conscious experience. Ground functionalism is built on the template for metaphysical explanation traced in §1, with mind making as the link.

---

14 It is also worth noting that the main evidence bearing on the examples *cross-cut* standard epistemic categories. For instance, the main evidence bearing on set formation is *a priori*, but the main evidence bearing on particle enactment is *a posteriori*, and the main evidence bearing on mereological fusion is a mix of *a priori* and *a posteriori* considerations. Insofar as inference to the best explanation is holistic, it should be expected that the full evidence for all principles involves every consideration there is, but that the main evidence for a given principle can be concentrated in various ways.

15 Hofweber (2009: 273) raises the concern that "esoteric" metaphysicians like myself are opening the door to *priority aquaism*, the view that everything is fundamentally water. Rabin (2019: 197) specifically charges that my views about consciousness lead to a "methodological madness" on which "anything goes," including *peanutism*, which is the view that "the actual world's fundamental level consists of Peanie the peanut." *Abductive methodology* provides my reply (see also Schaffer 2017: 14). Priority aquaism and peanutism look like explanatorily fruitless doctrines worth discarding. Note that I do not rule out such doctrines from the start. It may well have been reasonable for a counterpart of Thales to propose priority aquaism. I am instead saying that these doctrines now look fruitless and so deserve to be ignored, just as silly physical theories deserve to be ignored. Thus compare the silly metaphysical theory that what is fundamental is my nose, with the silly physical theory that what is causally initial is my nose ("the Big Sneeze"). I am saying that the right response in both cases is not to rule out such theories from the start for failing epistemic transparency, but rather to discard them now for seeming to be explanatorily fruitless. Here again the proper methodology is abductive.

2.1 Mind making

The ground functionalist posits a mind making principle, mapping realizer states to mental states by scanning the functional role of the input and molding a fitting output. The input to mind making is any token state of any system whatsoever, whether the system be an actual material system or some other-worldly ectoplasmic system. For instance, the input could be the current C-fiber firings (state) of Huma the human (system). This input is encased in the question, *what role does the state realize in the system*? For instance, it might realize the pain role, the tickle role, or no relevant role in the system at all.

The output of mind making is a token mental state (if any) for that system. For instance it could be Huma's current pain. This output is encased in the question, *what if any mental state is the system in*? For instance, it might be a pain, it might be a tickle, or it might be no mental state at all.

Mind making then maps input to output, pairing realizing states in a system to fitting mental states for that system, so forming an answer cascade from the deeper question *what role does the state realize in the system?*, to the downstream question *what if any mental state is the system in*? Thus:

> *Ground functionalism*: There is a grounding principle of mind making, which links realizer states in a system to mental states for that system

Just as right making may be said to imbue acts with moral status (§1.1), so mind making is said to imbue systems with mental properties.[16]

Realizer states in a system *enact* mental states for that system, and the ground functionalist views enactment as a specific metaphysical determination relation and thus a form of grounding, positing that what is enacted is thereby created. Indeed mind making and particle enactment (§1.1) are both enactment principles. Enactment may be thought of as a "dance it into being" principle. Just as what it takes to make a ballet is to have dancers spinning in the right patterns, so generally what it takes to make enacted things is to have the enactors functioning in the right patterns. A properly choreographed system dances out a mind.

So we can think of mind making as a machine. Feed it a a state of a system (material or otherwise), and it spits out a mental state on the basis of the role of the input. So if the role of pain is to be causally intermediary between tissue damage and avoidance behavior, and if C-fiber firings realize this role for Huma the human, then mind making maps Huma's C-fiber firing to her pain:

The mind making machine

Huma's C-fiber firing ⟶ Huma's pain
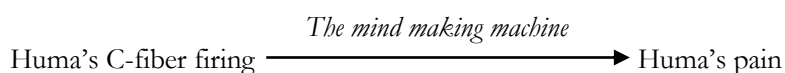
Likewise if what plays the pain role for Marvin the Martian is the inflation of small hydraulic cavities in his feet (as per the tale in Lewis 1980), then mind making maps Marvin's foot cavity inflation to his pain:

The mind making machine

[16] Some re-statements: For the Wilson-style pluralist who substitutes more specific metaphysical determination relations, just say: "There is a specific metaphysical determination relation of mind making…" For the Fine-style essentialist, say: "For each mental state, it lies in the nature of that state to be the state made by a state of a system realizing a fitting functional role." Though the essentialist will then make some counter-intuitive essence claims, for instance claiming that it lies in the nature of pain to be the state made by a state of a system caused by tissues damage and causing avoidance behavior. Intuitively, it simply lies in the nature of pain to feel a certain way. But insofar as the essentialist thinks of essences as real natures, it is not clear to me if she should or even can view our intuitions as reliable guides. (My preferred statement does not use the ideology of "essence" and so side-steps such issues.)

Marvin's foot inflation ——————————▶ Marvin's pain

Or if what plays the pain role for Caspar the ghost is a vibration of his ectoplasm, then mind making maps Caspar's ectoplasmic vibration to his pain:

The mind making machine

Caspar's ectovibration ——————————▶ Caspar's pain

Note that these are toy examples, in at least three respects. First, the actual role of pain is far more complicated, and indeed talk of pain in humans may need to give way to talk of distinctive states of nociceptive, neuropathic, and psychogenic pain. For the typical roles of each differ: nociceptive pain is typically caused by tissue damage, but neuropathic pain is typically caused by nervous system disorder, while psychogenic pain is typically caused by emotional distress.

Secondly, philosophical legend notwithstanding, C-fibers are almost certainly not the realizers of (nociceptive) pain in humans. C-fibers and Aδ-fibers are our two types of nociceptors in the peripheral nervous system, signaling certain "threatening" physical stimuli over to the central nervous system. Our realizers of pain are almost certainly found, not in local peripheral nervous system activations, but in more holistic central system patterns these tend to trigger.[17]

Thirdly and most relevantly, the roles of mental states are almost certainly interwoven, and so functional determination is best viewed as holistic, from the whole system at a time to a troupe of states for that system. So the deeper structure is really more akin to the structure of the functionalist principle of particle enactment (§1.1):

The mind enactment machine

System state —————————————▶ Mental state1
          —————————————▶ Mental state2
          —————————————▶ Mental state3

From a holistic functionalist perspective, minds are *natural dynamical structures*. Minds are not quite like the mousetraps or Coke dispensers commonly invoked in the literature (e.g. Fodor 1981: 120), for those are *artifacts*, and artifact kinds are determined in part by the intentions of their artificer. Rather minds are more like *ecosystems*, holistic and dynamic structures of interacting components enacted in nature. For instance, what makes a given ecosystem feature *a nutrient cycle* is that the system dances out an abstract and multiply realizable loop, seen in the way water on earth precipitates down from the sky and then evaporates back up.

[17] As to the human neural correlate of pain, Garcia-Larrea & Bastuji (2018) propose a model in terms of the coordinated activation of a three-part "pain matrix," involving: (i) unconscious processing in the sensorimotor and limbic areas, (ii) immediate stimulus awareness due to joint activation of the sensorimotor and fronto-cingulate-parietal networks, followed by (iii) integration into declarative perception linked to memory, involving the posterior cingulate and medial temporal areas. On this view, we might speak of the coordinated activation of Huma's pain matrix.

But just as it is easier to work with locally valid derivative models in the causal case, so it is easier in the mental case, and I will largely ignore the holistic aspect of functionalism in what follows. It should be more than enough even to exhibit even a locally valid and derivative explanation for conscious experience.[18],[19]

I say that—on these simplifying assumptions—the firing of Huma's C-fibers plus the principle of mind making *explain* why Huma is in pain (with parallel stories for Marvin and Caspar, etc.) After all, given that Huma's C-fibers are firing and mind making is at work, it is no mere accident that Huma is in pain. For we have a grounding principle: mind making, which links Huma's C-fiber firing to her pain. And we have input: Huma's C-fibers are firing. Huma's pain is the output of this principle on this input, and so she is in pain. This explanation reveals patterns in Huma's psychology, provides recipes for wiggling her pains, and allows one to understand why she is in pain by revealing how her pain is made.

I then say that this is a viable metaphysical explanation, with mind making playing the role of *link* from Huma's C-fiber firing to her pain. If we just considered her C-fiber firing and her pain on their own, we would just see a bucket of events with no specified relationship (likewise for the occurrence facts concerning these events—that is just a bucket of facts). That is not enough for explanation. Here there is an evident explanatory gap! But I think that the right response is not to declare consciousness an extra fundamental ingredient in reality(!!), but rather to recall that we generally need to involve links to see any explanatory connections. Mind making turns the trick (by design), connecting Huma's C-fiber firing to her pain in the direction of input to output, licensing an explanatory arrow drawn from C-fiber firing to pain, and charting how the answer to the question of her material state flows into an answer to the question of her mental state.

[18] Structural equation models for holistic systems are highly non-trivial. What is needed are mappings, for each downstream mental state, from the system state to it (akin to the Bohmian wave function realist mapping, for each downstream particle location, from the world location to it). First, we can think of the whole system as occupying a point in a *flowchart space*, which is a state-space of possible cognitive architectures and activations (including a point for "not a cognitive architecture"). We would then model the location of the system in flowchart space via a single independent variable. Second, we can think of each mental state as occupying a point in a *determination space*, for all of the possible determinate states (including a "no such state" point). We would then add many dependent variables to represent each mental state. Finally, for each mental state, we need a link from the place of the whole system in flowchart space, to the place of that mental state in its determination space. So for each dependent variable, we need a function from the value of the one independent variable to the value of that dependent variable This yields an answer cascade for each mental state, from the place of the whole system in flowchart space to the determinate status of that mental state. I trust it is clear why I have simplified away from holism!

[19] I leave open what the most basic or *root principle* is. Perhaps holistic mind enactment is simply a root principle in the workings of the world, or perhaps it stems from some deeper and more general enactment principle. Just as we might consider more and less restricted mereological fusion machines (§1.1), so we might consider more and less restricted enactment machines—including a universal enactment machine—from which mind enactment (and particle enactment, etc.) might derive. This is relevant to assessing whether or not *Ground functionalism* is inflationary. If mind enactment is a root principle, then the ground functionalist is positing an additional root principle in nature. But if mind enactment derives from a more general enactment principle needed anyway, then there is no additional cost (just as there is no additional cost in saying that the molecules of chemistry are made by mereological fusion, if mereological fusion was needed anyway).

So mind making backs the explanatory arrow that *bridges* the explanatory gap, in just the way that grounding principles generally link metaphysical explanations (§1.1; see also Schaffer 2017). Mind making connects material states to mental states as input to output, allowing for viable metaphysical explanation.[20],[21]

Indeed, this case can be modeled by exactly the same "two binary variable" structure seen in the simple case of Suki throwing the rock through the window, and the formation of {Socrates} from Socrates. Repeating the model but re-labeling the variables in obvious ways:

Variables1: <*Independents1*={*C-fibers*}, *Dependents1*={*Pain*}, *Range1*={*C-fibers*→{0,1}, *Pain*→{0,1}}>
Functions1: {*Pain* <= *Identity*(*C-fibers*)}
Assignment1: {*C-fibers*=1}

This structure induces the following directed graph, fitting our earlier representation:

C-fibers   ——————————————▶ *Pain*
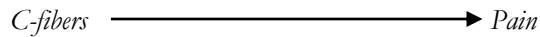
And moreover it allows us to solve for Huma's mental state (she is experiencing pain) from her material state.

In §1, when considering metaphysical explanation without yet looking at consciousness, I offered the following theses, repeated here:

Minimal explanatory claim: There are viable metaphysical explanations linked by grounding principles

Explanatory template: If the result is the output of the sources via the link, then there is a viable explanation of the result from the sources via the link

Template modeled: If there is an apt structural equation model in which the value of the variable representing the result is the output of the values of the variables representing the sources on a given function representing the link, then there is a viable explanation of result from sources and link

So I conclude that the ground functionalist style of explanation of Huma's pain from her C-fiber firings is a viable metaphysical explanation, fitting the template for metaphysical explanation linked by grounding principles, and thus exhibiting a viable explanation for consciousness.

Note that I only call this a viable explanation (successful given its background assumptions). I have not argued that this is the actual explanation, and indeed—as noted above—I regard the assumptions in play as simplified in multiple ways. But the explanatory gap arguments were supposed to show an in-principle

---

[20] In this vein, Tye (2000: 22) characterizes the epistemic gap in terms of a missing link:
> One who has a complete understanding of the term 'pain', for example, and who is fully apprised of the physical facts as we now know them, can still coherently ask why such-and-such brain processes or functional states feel the way pains do or why these processes feel any way at all. In this case, it seems that as far as our understanding goes, *something important is missing*. Herein lies the famous "explanatory gap" for consciousness.

I am saying that mind making adds what was missing.

[21] So what premise of Chalmers's (1996: 94–9) *zombie argument* am I rejecting? I am rejecting the premise that conceptual possibility entails metaphysical possibility. In general, scenarios that conflict with actual grounding principles may be conceptually but not metaphysically possible (for more detailed discussion see Schaffer 2017: 16–18). Since the zombie scenario violates the mind making principle, the ground functionalist has a principled reason to say that it is conceptually but not metaphysically possible. (Compare: Since a mereological nihilist scenario violates universal composition, the friend of universal composition has a principled reason to say that a mereological nihilist scenario is conceptually but not metaphysically possible.)

divide across which not even a viable explanation was possible. So I conclude that, whatever problems materialism might face, lack of a viable explanation for conscious experience is not among them.

2.2 Functionalism meets grounding
To further explicate ground functionalism, and situate it against some extant views, it is useful to decompose it into *functionalist* and *ground-theoretic* aspects. The functionalist aspect begins from an empirical claim of type-level *correlation*, between mental state types and second-order functional role properties, as per:

> *Functional correlation*: Every mental state type actually corresponds one-one to a functional role

For instance, the type pain—on the simplifications we are adopting—corresponds to the second-order functional role property of *being a property caused by tissue damage and causing avoidance behavior*.22

The primary rationale for *Functional correlation* is that—as a a matter of empirical fact—there seem to be robust correlations between material states and mental states, and that—given that states like pain seem to be multiply realizable across a wide range of diverse creatures (Putnam 1975)—these correlations seem best characterized at the functional level. In this vein Fodor (1981: 117–18) says that there is "a level of abstraction at which the generalization of psychology are most naturally pitched…" which "cuts across differences in the physical composition of the systems to which psychological generalizations apply."

Note that *Functional correlation* merely claims an *actual* correspondence. It is logically compatible with the incredible claim that this correspondence is just a cosmic accident. And it is compatible with the more credible claim that this correspondence holds for nomological reasons, in a way defended by *naturalistic dualists* like Chalmers (1996: 247; see also Gertler *forthcoming*):

> A natural suggestion is that consciousness arises in virtue of *the functional organization* of the brain. On this view, the chemical and indeed the quantum substrate of the brain is irrelevant to the production of consciousness. What matters is the brain's abstract causal organization, an organization that might be realized in many different physical substrates.

So *Functional correlation* seems to be an empirical truth, and is fairly widely accepted in the literature, on both the materialist and dualist sides of the aisle.

In my view, *Functional correlation* is the core insight of functionalism worth preserving (the other needed ingredient is grounding). It is because I accept *Functional correlation* that I am proposing *Ground functionalism*, and viewing the sources of the mind in terms of functional role.

But *Functional correlation* is not essential to my claim to deliver a materialist-friendly explanation of conscious experience. Indeed one of the main concerns about *Functional correlation* is Block's (1978) objection from overgeneration. For instance, if the population of China were to enact a system with the right roles, then *Functional correlation* would implausibly pair the population of China with mental states. My own

---

22 There is an intramural dispute among functionalists, as to whether to draw the mental state roles from folk psychology (Armstrong, Lewis) or from empirical science (Fodor). (The latter view, with its eye on psychology, is labeled "psycho-functionalism" in Block 1978.) *Functional correlation* is intended to be neutral, but insofar as it is said to represent an empirical truth, it seems to me that the right place to find causal roles is in empirical science. (Analytic functionalists like Lewis (1972) treat *Functional correlation* not as an empirical discovery but as part of folk theory, with mental state terms defined by Ramsification. Since my *Ground functionalist* is not making an analytic claim about the meaning of mental state terms, she need not follow.)

preference is to bite the bullet, but this is not a matter I can discuss here.23 But the main alternative is to say that there is a further requirement of the right kind of biological substrate, as per:

> *Functional-and-biological correlation*: Every mental state type actually corresponds one-one to a functional role realized in the right biological substrate

Or one might even drop the functional element altogether:

> *Biological correlation*: Every mental state type actually corresponds one-one to a biological substrate

The reader who prefers one of these alternatives is invited to replace the functionalist aspect in *Ground functionalism* accordingly, for instance via:

> *Ground functionalism-and-biologicism*: There is a grounding principle of mind making, which links realizer states in a system with the right biological substrate to mental states for that system

What bridges the explanatory gap is not taking a functionalist component as the relevant input, but positing a ground-theoretic mapping from any sort of material input to mental output.

Since *Functional correlation* is merely a claim of actual correlation, compatible with the incredible view that this is a vast cosmic accident, it is standard to add that the correlation has a basis, and that the mental thereby "arises from," "depends on," or "is determined by" the material (Fodor 1981: 118; Levin 2018).24 What is distinctive about *Ground functionalism*, comprising the ground-theoretic aspect of the view, is the follow-up claim that the basis of the correspondence is a metaphysical grounding principle, as per:

> *Metaphysical basis*: *Functional correlation* holds due to a metaphysical principle

The particular understanding of "arising" is then ground-theoretic:

> *Metaphysical arising*: Token mental states are grounded in token material states

This is naturally paired with the thought that these correlations are resilient to the extent that metaphysical grounding principles are resilient, which I will label *metaphysical necessity*, understood via the sphere of scenarios consistent with the grounding principles:

> *Metaphysical resilience*: *Functional correlation* holds with metaphysical necessity

Just as *nomological necessity* may be understood via the scenarios consistent with the laws of nature, so I understand metaphysical necessity via the scenarios consistent with "the laws of metaphysics".25

---

23 See Chalmers (1996: 249–51) for further discussion. I would just add that, in a literature in which the panpsychist attribution of mental states to all entities is considered a serious option, the functionalist attribution of mental states to a few surprising systems seems practically common-sensical in comparison!
24 Kriegel (*forthcoming-b*) notes that all parties can accept the general claim that the mental depends on the material, since the main dividing line "concerns the type of dependence involved. For the materialist, consciousness depends ontologically, metaphysically, or constitutively, on brain activity. For the dualist, the dependence is merely 'natural,' nomological, or causal."
25 According to Rosen (2006: 35), the laws of metaphysics "specify the categories of basic constituents and the rules for their combination. They determine how non-basic entities are generated from or 'grounded' in the basic array." Rosen says that the philosopher's conception of metaphysical necessity is indeterminate between a kind of necessity *tout court*, and a necessity restricted by the laws of metaphysics. I may be

I am now in position to situate *Ground functionalism* against extant dualist and functionalist views, in terms of the basis for and the resilience of *Functional correlation*. Let me start with *Naturalistic dualism*, which is the view that the material-mental correlations—which I take to be as given by *Functional correlation*—hold due to a merely contingent nomological connection. Thus Chalmers (1996: 125) clarifies: "[C]onsciousness arises from a physical substrate in virtue of certain contingent laws of nature…" And Kriegel (*forthcoming-a*) says:

> [Dualism] is the view (roughly) that the experiential and the physical are mutually (metaphysically) independent, such that any links between consciousness and its neural correlate are at most causal and contingent, not constitutive and necessary.

So working with *Functional correlation*, this yields:

> *Nomological basis*: *Functional correlation* holds due to a law of nature
> *Nomological arising*: Token mental states are caused by token material states

This is naturally paired with the thought that these correlations are resilient to the extent that laws of nature are resilient, which is *nomological necessity*, understood via the scenarios consistent with the laws of nature:

> *Nomological resilience*: *Functional correlation* holds with nomological necessity

So one way to reach *Ground functionalism* is to start from *Naturalistic dualism* plus *Functional correlation*, but strengthen the connections from nomological up to metaphysical.

Turning to analytic functionalism, this view builds in *Functional correlation* (indeed such is the main achievement of the view). But it views the correlation *conceptually*, in terms of the meanings of mental state terms. As Armstrong (1968: 82) says: "The concept of a mental state is primarily the concept of a state of the person apt for bringing about a certain sort of behaviour…" Thus the analytic functionalist holds:

> *Conceptual basis*: *Functional correlation* holds because terms for mental state types mean states with corresponding functional roles
> *Conceptual arising*: Token mental states are identical to token material states.

This is naturally paired with the thought that these correlations are resilient in any conceivable scenario (whether or not that scenario is consistent with the actual laws of nature, or the actual grounding principles):

> *Conceptual resilience*: *Functional correlation* holds with conceptual necessity
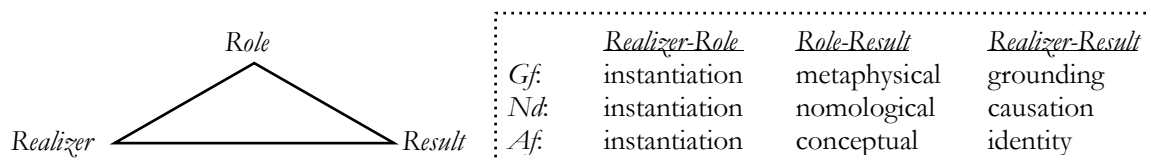
So a second way to reach *Ground functionalism* is to start from *Analytic functionalism*, but weaken the connections from conceptual down to metaphysical.[26]

As a related way to distinguish theses views, let us use the notions of *realizer* (e.g. C-fiber firings), *role* (e.g. causing avoidance), and *result* (e.g. pain). *Ground functionalism*, *Naturalistic dualism*, and *Analytic functionalism* agree that the *realizer-role* relation is one of instantiation. But the views disagree on whether the *role-result*

---

understood as adopting Rosen's restricted notion. Verbal matters aside, the key point is that there is this intermediate modal sphere to consider, which is neither merely nomological nor full conceptual necessity.
[26] *Ground functionalism* foregoes a functionalist analysis of mental state terms, and so falls silent on their meaning (as does *Naturalistic dualism*). *Ground functionalism* is thus consistent with a pure disquotational view, on which the only thing to be said about the meaning of a given mental state term '*t*' is that it means *t* (e.g. 'pain' means pain). In my view, there are few terms that allow for non-disquotational analyses, so there should be little expectation of a non-disquotational analysis of 'pain' or other mental state terms.

relation is a nomological, metaphysical, or conceptual connection, and on whether the *realizer-result* relation is causation, grounding, or identity. We can depict this via:

Role

Realizer ————————————— Result

	Realizer-Role	Role-Result	Realizer-Result
Gf:	instantiation	metaphysical	grounding
Nd:	instantiation	nomological	causation
Af:	instantiation	conceptual	identity

I am not claiming these three views to be exhaustive, but merely distinguishing between them. Other views worth considering include a a broadly Kripkean *functional identity* view (see for instance Papineau 1982 and Loar 1990: 99–101), which upholds the idea that the *Realizer-Result* relation is identity, but treats this identity not as a matter of *a priori* conceptual analysis, but rather as an *a posteriori* discovered identity in nature (like the water-H2O identity).[27]

Note that, in distinguishing these views, I am helping myself to the ideology of "nomological," "metaphysical," and "conceptual," as intuitive and distinct notions. (I treat these notions as correlated with different modalities, but would not explicate the distinctions modally—to my mind the modal differences are better viewed as manifestations of a fundamental distinction between *cause*, *ground*, and *definition*.) I trust that the distinctions I draw can be recognized, even by readers who prefer different terms. But the reader who will not tolerate my ideology must fend for herself.

2.3 The case for ground functionalism
I have introduced *Ground functionalism*, argued that it enables a viable explanation for conscious experience (§2.1), and distinguished it from extant views (§2.2). I conclude by arguing that *Ground functionalism* preserves the core advances of functionalism, is favored by *Abductive metaphysics* (§1.3) for being explanatorily fruitful, and is preferable to alternative views for this and other reasons.

As to preserving the core advances of functionalism, I take these to consist in reconciling materialism with inner mental life and multiple realizability. Working through these in reverse, it should be obvious that *Ground functionalism*—and indeed any view upholding *Functional correlation*—allows for multiple realizability. This is already exhibited in §2.1 with Huma the human, Marvin the Martian, and Caspar the ghost. *Ground functionalism* allows these diverse realizations to ground the same mental state, just by taking the grounding condition for a mental state to be the relatively abstract condition of realizing a functional role.

Moreover *Ground functionalism* allows for inner mental life. This is also exhibited with Huma, Marvin, and Caspar, where *Ground functionalism* allows us to attribute not just outward dispositions to behave (e.g. withdrawal), but inner conscious experiences (e.g. pain).

As to materialism, *Ground functionalism* preserves the materialist-friendly aspect of *Analytic functionalism*, by allowing that a purely material system can—in virtue of its functioning—enact mental states. Indeed, *Ground functionalism* is designed so as to entail that, if the actual realizers are material (as is empirically

---

[27] Chalmers (1996: 161–68) offers some more full-blown taxonomies, under which I take *Ground functionalism* to count as a form of "nonreductive" and "type B" materialism. But Chalmers does not use the ideology of grounding (to be fair, he was writing before grounding came onto center stage in metaphysics), and so neither distinguishes grounding-based from other versions of type B materialisms, nor anticipates how *Ground functionalism* fits into a general and independently motivated view of metaphysical explanation that emerged in the grounding literature. Indeed his (1996: 167) central resistance to Type B materialism is that "it relies on a metaphysics that is either incoherent or obscure, and one that is largely unmotivated; the main motivation is simply to avoid dualism at all costs." I may be understood as responding that grounding provides exactly the sort of general and motivated framework that the type B materialist needs.

plausible), then every actual mental state is grounded in the material. This result fits Loewer's (2001: 39) characterization of physicalism as "the fundamental properties and facts are physical and everything else obtains in virtue of them," fits Trogdon's (2013: 471; see also Bennet 2011: 33 and Dasgupa 2014: 557–59) "standard gloss of physicalism" as: "in the actual world each mental fact is ultimately grounded in certain physical facts…", and fits the view I elsewhere (Schaffer 2017: 14) label "*Ground physicalism*" on which "The physical is the ultimate ground for the chemical, the biological, and the psychological." In a related context Rosen (2010: 118; see also Pautz 2017: 388–89) says:

> [T]here is a difference between the materialist who holds that the facts about phenomenal consciousness are grounded in, and hence necessitated by, the neurophysiological facts that underline them, and the dualist for whom the neural facts merely cause or generate conscious states according to contingent causal laws.

Overall *Ground functionalism* is compatible with the world being fundamentally material (in the sense that all independent inputs into the grounding machinery are material states). It is also compatible with the idea that the material explains the mental (§2.3), that there is "nothing special" about the mental (like all higher-level phenomena, it is an output from fundamental physical inputs via the grounding machinery), and that the physical forms a metaphysical supervenience basis for the mental (where "metaphysical supervenience," as with "metaphysical necessity" above, holds grounding principles fixed). I say these are more than enough to qualify *Ground functionalism* as materialist-friendly (see also Schaffer 2017: 19–21).

That said, "materialism" is something of a cluster concept, and there are at least two theses in the cluster that *Ground functionalism* does not uphold. First, some materialists seek to completely excise mental terms from "the book of the world" (to borrow a phrase from Sider 2011). But the ground functionalist posits a mind making principle linking functional role (itself not a fundamental aspect of the material, but never mind that) to mental state, and so uses mental terms to describe the derivative outputs.[28]

Secondly, some materialists hold that every state of the world must be token-identical to a material state.[29] But the ground functionalist says that a given token mental state is grounded in the material, and given that grounding is *irreflexive*, it follows that the token mental state is not identical to, but merely dependent upon, its material grounds. (Nor is there any other material token to plausibly identify with the mental token.) Note that the ground functionalist is not saying that the mental state is wholly distinct from the material state either. There is an intermediate status between *identity* and *distinctness*, seen paradigmatically in the relation between mereologically partially overlapping things, and seen more generally—I say—between ground and derivative (Schaffer 2016: 75–6). I am saying that, on *Ground functionalism*, token mental states are neither identical to nor distinct from token material states, but have the intermediate status of the derivative.

I take these two points to show that there are recognizably materialist approaches that neither erase mental state terms from the book of the world, nor claim any token identities between material and mental states. But I do not care to squabble over labels. The reader is welcome to label *Ground functionalism* a form of "dualism" or "emergentism" (or perhaps a new position entirely), so long as she recognizes that *Ground functionalism* is built around the thesis that the mental is not fundamental but rather grounded in the physical. Indeed we may leave the cluster concept of materialism behind, and just consider this grounding thesis in its

---

[28] Given that grounding relations are generally backed by principles connecting the more to the less fundamental, principles generally will need higher-level terminology to characterize the non-fundamental outputs. There is nothing special about the mental here. For instance, if we want to recognize chemicals like H_2O molecules as derivative but real entities, then we will need principles that use the chemical terminology of "H_2O molecule" on the output side. No chemical terms on the output side, no chemical output.

[29] In this vein Nagel (1974: 446–47) says that "the meaning of physicalism is clear enough… mental events are physical events." But he adds that "when the two terms of the identification are very disparate it may not be so clear how it could be true" so that "an air of mystery surrounds the identification."

own right, as a basis for the empirical material-mental correlations seen in nature. The reader who thinks this grounding thesis is too weak is invited to show the evidence for a stronger claim.

This concludes my case that *Ground functionalism* preserves the core advances of functionalism. Given that *Ground functionalism* both enables a viable explanation for consciousness and preserves the core advances of functionalism, I add that is favored by *Abductive metaphysics* for being explanatorily fruitful. Or at least, given that we want a materialist metaphysics and an explanation for conscious experience (neither of which claim I have argued for here, but both of which strike me as deeply plausible), a mind making principle is worth the positing. So I say that *Ground functionalism* provides a new chapter in the hopeful materialist story of the mind, preserving the insights of functionalism while integrating the developments of metaphysical grounding to achieve explanation.

I have focused on the positive development of *Ground functionalism*, and have not attempted to criticize other views. But I close with a parting glance at some comparative reasons for preferring *Ground functionalism* over *Naturalistic dualism, Analytic functionalism*, and functional identity theories, if only to trace some changes in the usual dialectic, and point to some lines of further discussion.

So first, as compared to *Naturalistic dualism, Ground functionalism* is mainly to be preferred for allowing materialism.[30] But a second reason to prefer *Ground functionalism* is that the material-mental correlations do not seem nomological (as the naturalistic dualist claims), in at least three respects. For the most fundamental nomological connections of our world seem to have the following character:

- They link *fundamental* states of nature
- They are *global*, concerning whole states of the cosmos
- They are *dynamic*, governing the temporal evolution of systems (or expressing global constraints)

In this vein Maudlin (2007: 172) speaks of "fundamental laws of temporal evolution" which "specify how the state of the universe will, or might, evolve from a given initial state."

Now consider a given psycho-physical connection, such as the C-fiber firing to pain link. I say that this link has none of the bulleted characteristics of a fundamental law. It does not link fundamental states, since it invokes C-fiber firings on the left, which are not fundamental on any account. (Note that I am not complaining that this connection invokes pain on the right, as that complaint would beg the question against the dualist; rather I am complaining that this connection invokes C-fiber firings on the left, which are not fundamental material states on any account.) It is not global (nor perfectly local) but rather regional, operating at a middle-sized biological scale. And it is not dynamic but synchronic. Such regional and synchronic ropes through the nonfundamental look like vertical grounding connections (as per *Ground functionalism*) rather than horizontal causal connections (as per *Naturalistic dualism*).[31]

Secondly, as compared to *Analytic functionalism, Ground functionalism* is mainly to be preferred for explaining consciousness. But another reason to prefer *Ground functionalism* is that *Analytic functionalism* is based on meaning claims that are just obviously false, and indeed are already refuted by the mere conceivability of zombie scenarios. For recall that analytic functionalism claims that it is conceptually necessary that realizing

---

[30] Thus dualists often acknowledge that they would prefer materialism but were only driven to dualism by the arguments. For instance, Chalmers (1996: xiv) says:

> For a number of years, I had hoped for a materialist theory; when I gave up on this hope, it was quite reluctantly. It eventually seemed plain to me that these conclusions were forced on anyone who wanted to take consciousness seriously. Materialism is a beautiful and compelling view of the world, but to account for consciousness, we have to go beyond the resources it provides.

I am saying that *Ground functionalism* offers a new hope for materialism.

[31] See Schaffer (*forthcoming*) for further criticism of the laws of nature posited by the naturalistic dualist.

the functional role for a mental state means having that mental state. But—given that the zombie scenario is conceptually possible—it is conceptually possible to have a functional duplicate of our world but which is devoid of any conscious experience at all. So it immediately follows that analytic functionalist's claim of conceptual necessity is false, *before any questionable jump from conceptual necessity to metaphysical necessity.* Hence I agree with Chalmers (1996: 17–18), when he claims: "It is not a *conceptual* truth that the process should be accompanied by the phenomenal quality, but it is a fact about the world." I am just proposing that this fact about the world is not a contingent nomological connection as Chalmers then infers, but something intermediate he did not consider: a metaphysically necessary grounding connection.[32]

The third and final parting comparison I offer is with functional identity theories. Again *Ground functionalism* is mainly to be preferred for explaining consciousness. But a second reason to prefer *Ground functionalism* is that, while functional identity theories dispense with obviously false claims of conceptual analysis, they leave in their wake bare and incredible identity assertions. For instance, the property of *having a property caused by tissue damage and causing avoidance behavior* and the property of *being in pain* look like clearly different properties (with different extensions at zombie scenarios). There is far more to say. For instance perhaps this appearance of difference is just an illusion due to specific features of our phenomenal concepts (see Tye 1980, Loar 1990, and Papineau 1982 on "conceptual dualism"; but see Chalmers 2007 for an argument against such views). But for now suffice it to say that the ground functionalist avoids such bare and incredible identity assertions.

A third reason to prefer *Ground functionalism* is that functional identity theories have trouble reconciling the unity and the causal efficacy of mental properties (Kim 1998). There is a core dilemma between *role theories* which identify mental properties with causally inefficacious second-order properties, and *realizer theories* which identify mental properties with disunified first-order realizing properties. There is far more to say (see Robb & Heil 2018 for a useful overview of the myriad issues arising). But for now suffice it to say that the ground functionalist allows mental properties to stand as unified first-order properties.

Leaving this parting glance at comparisons aside, my primary claim remains that *Ground functionalism* achieves a viable materialist-friendly explanation for conscious experience. Nagel (1974: 445) says: "If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us a clue how this could be done." I claim to have shown how this can be done by positing that mind making is at work. Minds are made by dancing, and even matter can dance.[33]

---

[32] The functionalist literature is replete with examples of terms that really do seem to have functionalist meanings, like 'mousetrap'. But it seems to me that phenomenal terms just do not have these sorts of functionalist meanings. To the extent that we can say anything substantive about their meanings beyond bare disquotational claims like "'pain' means pain," these terms seem to name types of inner subjective feels, not causal abilities and liabilities. Small wonder that analytic functionalism was unable to properly explain inner experience, when it began by mis-defining it.

References

Albert, David 1996. Elementary Quantum Metaphysics. *Bohmian Mechanics and Quantum Theory: An Appraisal*, eds. James T. Cushing, Arthur Fine, and Sheldon Goldstein: 277–84. Dordrecht: Kluwer.

Allori, Valia 2013. Primitive Ontology and the Structure of Fundamental Physical Theories. *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*, eds. Alyssa Ney and David Albert: 58–75. Oxford: Oxford University Press.

Armstrong, David 1968. *A Materialist Theory of the Mind*. London: Routledge.

Bennett, Karen 2011. By Our Bootstraps. *Philosophical Perspectives* 25: 27–41.

Block, Ned 1978. Troubles with Functionalism. *Minnesota Studies in the Philosophy of Science* 9: 261–325.

Chalmers, David 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

—— 2007. Phenomenal Concepts and the Explanatory Gap. *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, eds. Torin Alter and Sven Walter: 167–94. Oxford: Oxford University Press.

—— 2012. *Constructing the World*. Oxford: Oxford University Press.

Chalmers, David and Frank Jackson 2001. Conceptual Analysis and Reductive Explanation. *The Philosophical Review* 110: 315–60.

Dasgupta, Shamik 2014. The Possibility of Physicalism. *The Journal of Philosophy* 111: 557–92.

Feynman, Richard 1965. *The Feynman Lectures on Physics, vol. 3: Quantum Mechanics*, eds. Richard Feynman, Robert Leighton, and Matthew Sands. Reading, MA: Addison-Wesley.

Fine, Kit 2001. The Question of Realism. *Philosophers' Imprint* 1: 1–30.

—— 2012. Guide to Ground. *Metaphysical Grounding: Understanding the Structure of Reality*, eds. Fabrice Correia and Benjamin Schnieder: 37–80. Cambridge: Cambridge University Press.

Fodor, Jerry 1981. The Mind/Body Problem. *Scientific American* 244: 114–23.

Garcia-Larrea, Luis and Hélène Bastuji 2018. Pain and Consciousness. *Progress in Neuro-Pharmacology and Biological Psychiatry* 87B: 193–99.

Gertler, Brie 2001. The Explanatory Gap is not an Illusion: Reply to Michael Tye. *Mind* 110: 689–94.

—— *forthcoming*. Dualism: How Epistemic Issues Drive Debates about the Ontology of Consciousness. *Oxford Handbook of the Philosophy of Consciousness*. Oxford: Oxford University Press.

Glazier, Martin 2020. Explanation. *Routledge Handbook of Metaphysical Grounding*, ed. Mike Raven: 121–32. London: Routledge.

Hofweber, Thomas 2009. Ambitious, yet Modest, Metaphysics. *Metametaphysics*, eds. David Chalmers, David Manley, and Ryan Wasserman: 260–89. Oxford: Oxford University Press.

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

Kim, Jaegwon 1998. Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. Cambridge, MA: The MIT Press.

Kosman, L. A. 1973. Understanding, Explanation, and Insight in the *Posterior Analytics. Exegesis and Argument: Studies in Greek Philosophy Presented to Gregory Vlastos*, eds. Edward Lee, Alexander Mourelatos, and Richard Rorty: 374–92. New York: Humanities Press.

Kriegel, Uriah *forthcoming-a*. What is the Philosophy of Consciousness? *Oxford Handbook of the Philosophy of Consciousness*. Oxford: Oxford University Press.

—— *forthcoming-b*. Beyond the Neural Correlates of Consciousness. *Oxford Handbook of the Philosophy of Consciousness*. Oxford: Oxford University Press.

Levin, Janet 2018. Functionalism. *Stanford Encyclopedia of Philosophy*: https://plato.stanford.edu/entries/functionalism/ (accessed 4/4/20).

Levine, Joseph 2001. *Purple Haze*. Oxford: Oxford University Press.

Lewis, David 1972. Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy* 50: 249–58.

—— 1980. Mad Pain and Martian Pain. *Readings in Philosophy of Psychology, Volume 1*, ed. Ned Block: 216–32. Cambridge, MA: Harvard University Press.

Lewis, David and Stephanie Lewis 1970. Holes. *Australasian Journal of Philosophy* 48: 206–12.

Loar, Brian 1990. Phenomenal States. *Philosophical Perspectives* 4: 81–108.

Maudlin, Tim 2007. *The Metaphysics within Physics*. Oxford: Oxford University Press.

—— 2010. Can the World Be Only Wavefunction? *Many Worlds? Everett, Quantum Theory, & Reality*, eds. Simon Saunders, Jonathan Barrett, Adrian Kent, and David Wallace: 121–43. Oxford: Oxford University Press.

Montero, Barbara 2013. Must Physicalism Imply the Supervenience of the Mental on the Physical? *Journal of Philosophy* 110: 93-110.

Nagel, Thomas 1974. What is it Like to Be a Bat? *The Philosophical Review* 83: 435–50.

Ney, Alyssa 2013. Ontological Reduction and the Wave Function Ontology. *The Wave Function: Essays on the Metaphysics of Quantum Mechanics*, eds. Alyssa Ney and David Albert: 168–83. Oxford: Oxford University Press.

Papineau, David 2002. *Thinking about Consciousness*. Oxford: Oxford University Press.

Paul, L. A. 2012. Metaphysics asd Modeling: the Handmaiden's Tale. *Philosophical Studies* 160: 1–29.

Pautz, Adam 2017. The Significance Argument for the Irreducibility of Consciousness. *Philosophical Perspectives* 31: 349–407.

Pearl, Judea 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Putnam, Hilary 1975 (orig. 1967). The Nature of Mental States. *Mind, Language, and Reality: Philosophical Papers Volume 2*: 429–40. Cambridge: Cambridge University Press.

Quine, W. V. O. 1948. On What There Is. *Review of Metaphysics* 2: 21–38.

Rabin, Gabriel 2019. Grounding the Gaps or Bumping the Rug? On Explanatory Gaps and Metaphysical Methodology. *Journal of Consciousness Studies* 26: 191–203.

Railton, Peter 1986. Moral Realism. *The Philosophical Review* 95: 163–207.

Robb, David and John Heil 2018. Mental Causation. *Stanford Encyclopedia of Philosophy*: https://plato.stanford.edu/entries/mental-causation/ (accessed 4/4/20).

Rosen, Gideon 2006. The Limits of Contingency. *Identity and Modality*, ed. Fraser MacBride: 13–39. Oxford: Oxford University Press.

—— 2010. Metaphysical Dependence: Grounding and Reduction. *Modality: Metaphysics, Logic, and Epistemology*, eds. Bob Hale and Aviv Hoffmann: 109–36. Oxford: Oxford University Press.

Schaffer, Jonathan 2009. On What Grounds What. *Metametaphysics*, eds. David Chalmers, David Manley, and Ryan Wasserman: 347–83. Oxford: Oxford University Press.

—— 2010. Monism: The Priority of the Whole. *The Philosophical Review* 119: 31-76.

—— 2016. Grounding in the Image of Causation. *Philosophical Studies* 173: 49–100.

—— 2017. The Ground between the Gaps. *Philosophers' Imprint* 17: 1–26.

—— 2018. Laws for Metaphysical Explanation. *Philosophical Issues* 27: 302–21.

—— *forthcoming*. Naturalistic Dualism and the Problem of Physical Correlates. *Grounding and Consciousness*, ed. Gabriel Rabin. Oxford: Oxford University Press.

Trogdon, Kelly 2013. Grounding: Necessary or Contingent? *Pacific Philosophical Quarterly* 94: 465–85.

Tye, Michael 2000. *Color, Consciousness, and Content*. Cambridge, MA: The MIT Press.

van Inwagen, Peter 1990. *Material Beings*. Ithaca, NY: Cornell University Press.

Wilsch, Tobias 2015. The Nomological Account of Ground. *Philosophical Studies* 172: 3293–3312.

Wilson, Jessica 2014. No Work for a Theory of Grounding. *Inquiry* 57: 535–79.

Woodward, James 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.