

# **Introverted Metaphysics**

## How We Get Our Grip on the Ultimate Nature of Objects, Properties, and Causation

Uriah Kriegel

*Metaphilosophy* 50 (2019)

**Abstract ::** This paper pulls together three debates fundamental in metaphysics and proposes a novel unified approach to them. The three debates are (i) between bundle theory and substrate theory about the nature of objects, (ii) dispositionalism and categorialism about the nature of properties, and (iii) regularity theory and production theory about the nature of causation. The first part of the paper suggests that although these debates are metaphysical, the considerations motivating competing approaches in each debate tend to be epistemological. The second part argues that the two underlying epistemological pictures supporting these competing views lead to highly unsatisfying conceptions of the world. The final part proposes an alternative epistemological picture, which I call “introverted empiricism,” and presents the way this alternative provides for a more satisfying grasp of the ultimate nature of objects, properties, and causation. It is a consequence of this alternative picture that there is a kind of intimate self-understanding that underlies our understanding of the deep nature of reality.

### **1. Introduction: The world around us and its fundamental patterns**

When I wake up in the morning, I find my wife beside me and our dog squeezed between us. She looks calm and he feels warm. The faint noise of their breathing envelops me while I ease into consciousness. As I step out of the blanket, the

coldness of the air hits me. I head to the kitchen and turn on the faucet so that water fills my glass. I can see and hear the burgundy clock ticking overhead, as I open the fridge, take out a yogurt, and use a silver spoon to swivel into it some chestnut honey. I take all this to my desk, turn on my computer, and start answering emails or working on the next paragraph of whatever I'm writing.

Up till that point, the patterns of my life are very stable – the first ten minutes of every day look essentially the same. From that point on, however, I cross the border into the uncontrollable and lawless region of life – every day looks completely different. The variety of elements that might float in and out of the theater of my life on any given day is immense. And yet, some stubborn patterns persist here too, though they are much more abstract patterns than warm dogs and chestnut honey. In particular, what populates my life at every turn are individual objects, their properties, and causal interactions among them. My wife and dog, the clock and the fridge, yogurt and computer – these are some of the first objects I encounter, but a great many follow. The properties of looking calm, feeling warm, being silver, burgundy-colored or chestnut-flavored that qualify these objects are followed by other properties qualifying other objects. Finally, enveloping, hitting, ticking, turning, swiveling – these causal transactions, in which I function as agent, patient, or just witness – give way to innumerable many others as the day progresses.

It's an open question whether a complete description of all objects, properties and causal relations in our universe would leave anything out of the great book of the world.<sup>1</sup> What's clear is that such a complete description would cover a substantial portion of our *Lebenswelt*, the world as we experience it around us. It's also clear that understanding the ultimate nature of objects, properties and causation is the province of philosophy. Different sciences study specific kinds of object: particles are studied by physics, molecules by chemistry, cells by biology, and so on. But no science studies *the object as such*. Likewise, my wife's property of being female is studied by biology while her property of being calm is studied by psychology, with many other properties of hers studied by physics, chemistry, sociology, economics and so on – but no science is in charge of studying *properties qua properties*. Similarly for causation: different causal laws are formulated by different sciences, but no science instructs us on

the nature of *causation itself*. It is the mandate of philosophy to address the nature of objects, properties and causation *in abstracto*.

Interestingly, philosophical debates over the nature of objects, properties and causation have tended to organize around two poles. Indeed, at the relevant level of abstraction, they are *the same* two poles. To this *philosophical* pattern, if you will, we now turn.

## 2. Objects: Property bundles and underlying substrata

The two poles of the traditional debate on the nature of objects are the *bundle theory* and the *substratum theory*. According to the bundle theory, an object is in reality just a bundle of properties. My dog Julius, for example, is nothing more than the sum of his properties: being warm-blooded, being four-legged, being pumpkin-colored, and so on. More generally:

(BT) For every object  $O$ , there is a collection of properties  $P_1, \dots, P_n$ , such that  $O = P_1, P_2, \dots, P_n$ .

According to the substratum theory, in contrast, an object is not just a bundle of properties, but rather a certain *je-ne-sais-quoi* that underlies, bears, or supports those properties. That underlying we-know-not-what is often labeled ‘substratum’:

(ST) For every object  $O$ , there is a substratum  $S$  and a collection of properties  $P_1, P_2, \dots, P_n$ , such that (i)  $S$  supports  $P_1, P_2, \dots, P_n$  and (ii)  $O = S$ .

On this view, Julius is not the sum of his properties, but the substratum that underlies and *has* those properties.

The arguments for and against BT and ST are immensely varied and highly sophisticated. But the basic *motivation* for each view is simple and straightforward. Interestingly, although the theories themselves are *metaphysical* and concern the *nature* of objects, the fundamental motivations for them are often *epistemological* and based on the kind of *insight* we may have into

objects. The basic motivation for the bundle theory, I contend, is an empiricist epistemology; that for the substratum theory, a rationalist epistemology.

For the empiricist, any insight we have into *anything* is ultimately based on perceptual experience. But when we observe objects, all we can perceive are their various properties. Therefore, the insight we have into an object is exhausted by the sum of its properties. Some properties of my dog I perceive *directly*, with my own senses: I can see that he's pumpkin-colored. Others I perceive only *indirectly*, with the aid of common measuring tools: I see that he weighs eleven kilos by seeing the needle on the scale settle on 11. Yet other properties are perceptually accessible only with rarer instruments, typically found only in scientific laboratories: Julius' DNA, for example, can be revealed by an autosomal DNA test. But for the empiricist, anything we can observe and record about Julius, however indirectly, is bound to be a property of some sort. We are never going to suddenly encounter something that is not itself a property but mysteriously underlies and supports properties. Thus Hume describes substratum as a *fiction*:

... the particular qualities [or properties], which form a substance [or object], are commonly referred to an unknown something, in which they are supposed to inhere [i.e., a substratum]; or granting this fiction should not take place, are at least supposed to be inseparably connected... (*A Treatise of Human Nature*, I.i.6)

No attentive examination of an object and no scientific study, however thorough, could reveal anything but further properties of the object. So to posit a substratum is to posit something of which we have no possible positive conception – a *fiction*, in Hume's words.

At the same time, insists the rationalist, the very notion of a property suggests something that *has* the property. It is strictly unintelligible to speak of a property that is not a property of *something*. Accordingly, it's hard to make sense of the notion that Julius is just a cloud of properties floating about, miraculously sticking together, without anything that *has* those properties and underlies their togetherness. For the rationalist, this is why we must posit a substratum: it's the only way to *render intelligible* the object. In other words, the substratum is a transcendental condition on the very intelligibility of the bundles of properties we perceive. Kant writes:

[A] transcendental substratum grounds the complete determination of things in our reason – a substratum which is to form the stock from which all possible predicates of things [properties of objects] are taken... It is thus a transcendental ideal which ... is the complete and highest substantive condition of [the object's] possibility, to which condition must be traced back the thinking of all objects... (*Critique of Pure Reason*, A 575-6/B 603-4)

In other words, the thing in itself is the substratum without which objects are unintelligible. Kant can agree with Hume that the only perceptually accessible aspect of an object is its properties. But he denies that our insight into something is limited to what is perceptually accessible. On the contrary, sometimes reason finds itself forced to posit something which in itself is in principle imperceptible, but which renders intelligible that which is perceivable. To make sense of the bundles we perceive, we must posit an underlying substratum that supports them and accounts for their bundling.

Hume would reject this, naturally, claiming that since we can form no positive conception of this ultimate substrate, we do not really have any insight into its nature. At bottom, then, for the empiricist an object is something essentially perceptible, whose nature must therefore be itself perceptually manifest; whereas the rationalist emphasizes the *intelligibility* of the object, and accordingly construes its essence as accessible primarily to the intellect, not sense perception.

### 3. Properties: Causal dispositions and underlying categorical bases

Historically, the main debate about the nature of properties has been between platonism and nominalism. In current metaphysics, however, the more prominent debate is between dispositionalism and categorialism. To understand this debate, we must appreciate the distinction between a disposition and its categorical basis. Explosiveness is a disposition: something can be explosive even if it never explodes, provided that it is suitably disposed to explode. What this means is itself up for debate, but in any case, a disposition is always a disposition to have a certain causal effect; it is, we may say, a cluster

of causal powers. Many metaphysicians hold that where we find such a cluster of causal dispositions, we should expect them to be grounded in categorical bases: some non-dispositional properties that underlie and *explain* the disposition. A bomb is explosive because it contains potassium nitrate. The property of containing potassium nitrate then grounds and explains the bomb's explosiveness. It is the *categorical basis* of the explosiveness.

With this distinction in place, we can formulate dispositionalism and categoricalism with some precision. For reasons that we will see momentarily, dispositionalists claim that in fact *all* properties are dispositional. Categoricalists claim that at least some properties are categorical, namely, the *fundamental* ones. (For present purposes, we may understand fundamental properties as properties of fundamental physical particles.) According to dispositionalism, all such fundamental properties are dispositional:

- (D) For every fundamental property F, there is a collection of causal powers  $P_1, \dots, P_n$ , such that  $F = P_1, P_2, \dots, P_n$ .

According to categoricalism, there may be dispositional properties at higher levels of reality, but fundamental properties are categorical one and all:

- (C) For every fundamental property F, there is a categorical property C and a collection of causal powers  $P_1, P_2, \dots, P_n$ , such that (i) C is the categorical basis of  $P_1, P_2, \dots, P_n$  and (ii)  $F = C$ .

As before, the motivation for D and C is essentially epistemological and pertains to the kind of insight we can have into the nature of properties. The empiricist claims that empirical science characterizes fundamental properties only in terms of the causal laws that govern them and does not comment on any categorical "quiddity" that might underlie this lawful behavior. The rationalist, in contrast, insists that a cluster of dispositions to behave in certain ways is unintelligible without the supposition of such a categorical quiddity underlying and explaining it.

Some philosophers have argued that on closer examination properties such as "containing potassium nitrate" are themselves dispositional, though less obviously so than properties such as explosiveness. Potassium nitrate is a

compound of potassium and nitrate; nitrate itself is a compound of one nitrogen atom and three oxygen atoms; and so on. But what do we know about nitrogen? What makes an atom a nitrogen atom? All empirical science tells us is that nitrogen is the kind of atom sufficiently large collections of which are solid below  $-210^{\circ}\text{C}$ , liquid between  $-210^{\circ}\text{C}$  and  $-195^{\circ}\text{C}$ , and gaseous above  $-195^{\circ}\text{C}$ . Ultimately, these are just dispositional properties: a gaseous collection of nitrogen atoms is *disposed* to turn liquid at  $-195^{\circ}\text{C}$  and *disposed* to turn solid at  $-210^{\circ}\text{C}$ , even if these dispositions are never manifested; a solid collection of nitrogen atoms *would* melt at  $-210^{\circ}\text{C}$  and *would* boil at  $-195^{\circ}\text{C}$ , even if it never does melt or boil. For an atom to have the property of being nitrogen, then, is for it to have the right cluster of (causal) dispositions. More generally, atoms as well as subatomic particles are all empirically characterized by science in terms of their distinctive funds of causal powers, that is, their dispositions to enter into causal processes described by scientific laws.

On this basis, some philosophers have adopted the view of “dispositions all the way down”. Consider Karl Popper:

The view of propensities [i.e., dispositions] allows us to see in a new light the processes that constitute our world: the world process. The world ... can now be seen as a world of propensities, as an unfolding process of realizing possibilities [or manifesting dispositions] and of unfolding new possibilities. (*A World of Propensities*, pp. 18-9)

The motivation for this view, to repeat, is that we cannot empirically establish anything about fundamental properties beyond how they *would* causally impact their environment under such-and-such conditions.

For the categoricalist, however, there is something essentially unintelligible about the idea of dispositions all the way down. Suppose a particle has fundamental property  $F_i$  just if it is disposed to cause (in the right circumstances) certain effects, such as another particle acquiring  $F_j$  and a third one losing  $F_k$ . This illuminates the nature of  $F_i$  only if we already know what  $F_j$  and  $F_k$  are. In characterizing  $F_j$  and  $F_k$ , however, the dispositionalist only mentions further potential effects, involving  $F_l$  and  $F_m$ . But when does the system actually go beyond further dispositions and actually *realize* some potentials? As David Armstrong puts it, “propensities all the way down” means “always packing and never going” (*A World of States of Affairs*, p. 80) – always ready to

do something, but without anything ever actually happening. Eventually, something must actually *happen* in the world, reasons the categoricalist, so there must be more to (some) properties than dispositions, potentials and powers. They must include a categorical basis of dispositions, a ground for those powers and potencies. Accordingly, categoricalism posits an intrinsic *je-ne-sais-quoi* in every fundamental property that serves as categorical basis for the property's fund of causal dispositions – what is sometimes referred to as the property's "quiddity."

According to Rae Langton's "Kantian Humility" thesis, these categorical properties are nothing but Kant's noumenon – something whose nature we should humbly concede not to know, but which we must posit nonetheless. She writes:

[If] the [categorical] ground is distinct from the [causal] power, and contingently connected with it, then our orthodoxy is faced with a conclusion surprisingly similar to Kantian Humility... [Our name for the ground] becomes the name for a something-we-know-not-what – ominously similar to a Kantian thing in itself. (*Kantian Humility*, p. 176)

The categorical properties are empirically intractable – science offers us no insight into their nature. Our only reason to believe in them is transcendental: if they did not exist, nothing would *actually happen* in the world. When a particle has mass  $m$ , for instance, it may well become disposed to affect other particles in certain ways; but in addition, something entirely non-dispositional occurs: the intrinsic categorical  $m$ -ness is instantiated. These categorical properties are thus the "meat" of the world; without them, all we have is what Russell called "the causal skeleton of the world" (*The Analysis of Matter*, p. 391).

In summary, as in the case of objects, this debate over the metaphysics of properties seems fueled by an epistemological contrast between empiricism and rationalism: dispositionalism is motivated by considerations of *empirical tractability*, categoricalism by considerations of *rational intelligibility*. The ultimate source of the dispositionalist conception of properties is empirical inquiry, that of the categoricalist conception transcendental reasoning.

#### 4. Causation: Regularities and the underlying 'secret connexion'

The central historical debate on the nature of causation is between *regularity theories* and *production theories*. What does the fact that fire causes heat consist in? According to regularity theories, it consists in the fact that whenever there is fire, heat soon follows; the two appear one before the other with consistent regularity. According to production theories, this regularity between fire and heat is but a *symptom* of a deeper connection between them that *underlies* and *explains* the regularity; causation proper is that underlying ‘secret connexion’ (as Hume called it) whereby the cause actually *produces* or generates the effect.

To express the two theories more precisely, observe that causal claims can be made both about *types* and about *tokens*: “smoking causes cancer” is a claim about causal relation between types; “the arsonist caused the fire” is a claim about causal relation between tokens. In other words, we must distinguish two causal relations: type-causation and token-causation. Regularity theory says that type-causation consists simply in a regularity relation between the relevant types of event; token-causation occurs when there is a regularity relation between the relevant event-types and one token event precedes the other. More precisely:

- (RT) For every token events  $a$  and  $b$ ,  $a$  causes  $b$  iff there are event-types A and B, such that (i)  $a$  is a token of A, (ii)  $b$  is a token of B, (iii) there is a regularity relation between A and B, and (iv)  $a$  precedes  $b$ .

The production theory, meanwhile, says that type-causation occurs when the secret connection underlying regularity holds between the relevant types’ tokens. As regards token-causation, the theory is succinct if uninformative:

- (PT) There is a secret connection C, such that for every token events  $a$  and  $b$ ,  $a$  causes  $b$  iff  $a$  bears C to  $b$ .

This secret connection may be labeled ‘production’, but this is just a convenient label. PT does not tell us what the secret connection is.<sup>2</sup>

RT and PT are metaphysical theories about the *nature* of the causal relation. But the main motivations for them, again, are epistemological and

concern the kind of insight we can have into the causal relation. Clearly, Hume supported RT on empiricist grounds:

[Perceptual] experience only teaches us, how one event constantly follows another; without instructing us in the secret connexion, which binds them together, and renders them inseparable. (*An Enquiry Concerning Human Understanding*, §52)

Whenever we perceive fire, we soon perceive heat as well. We thus perceive the “constant conjunction” between them. But we never see the fire actually producing the heat – that connection between them remains imperceptible, “secret”. A good empiricist must therefore think of the world as a vast sequence of perceptible events, with certain perceptible patterns of regular co-occurrence between them – and nothing more.

Granted a constant conjunction between fire and heat, claims the rationalist, we are still tempted to ask *why* there is such a conjunction. It seems completely implausible to say that the conjunction is “one big coincidence”, an inexplicable fact for which there is no *reason*. Surely there is no miracle here; rather, heat follows fire with regularity precisely because something in the fire produces the heat. Now, the regularity theorist can insist that the fire’s production of heat merely consists in a series of empirically discoverable exothermic chemical reactions, each of which is a causal exchange itself consisting merely in the regularity between the cause and the effect. This account must bottom out, however, in microphysical causal relations at the fundamental level of reality. Suppose now that at this fundamental level event  $E_1$  causes event  $E_2$ . For the regularity theorist, all this means is that  $E_1$  and  $E_2$  entertain regular “constant conjunction”. Here too, however, we may want to know *why* they do, what ensures that whenever  $E_1$  is tokened  $E_2$  is immediately tokened as well. The production theorist offers an answer: because  $E_1$  tokens produce  $E_2$  tokens, where the production of the latter by the former is an imperceptible, empirically intractable, ‘secret’ connection. The rationalist insists that we *must* posit such an empirically intractable connection to *explain* and make sense of the empirically established regularities. Without the underlying secret connection, regularity relations are strictly miraculous. As before, then, we need this secret connection, even though we have never laid eyes on it, to render *intelligible* what our eyes report to us.

The only reason to believe in the secret connection between cause and effect, then, is *transcendental*: the very possibility of non-miraculous regularity requires an underlying connection between cause and effect. For the rationalist, it is true that we have no *perceptual* insight into this underlying connection, but we have an *intellectual* insight into it inasmuch as we grasp the kind of role it plays in rendering intelligible the perceptible regularities. The empiricist insists, however, that this is no real insight, since the rationalist admits we are unable to form any *positive conception* of the secret connection (that's why it is *secret!*). All we know is what the secret connection is *not*, and what kind of explanatory *role* it must play. But the intrinsic *nature* of that which plays the role, being something we have never actually witnessed, remains unknown and unknowable. That hardly qualifies as "insight".

## 5. Humean empiricism and the phenomenal skeleton of the world

More generally, for objects, properties and causation alike, traditional metaphysical disputes seem epistemologically anchored. At bottom, our conception of the *nature* of objects, properties and causal relations depends on the kind of insight we seek into them. The empiricist seeks insight into the way objects, properties, and causation perceptually appear to us. The rationalist seeks insight into the underlying *je-ne-sais-quoi* that constitutes, according to her, the real essence of objects, properties, and causation.

The Humean empiricist faces an immediate problem, however. She refuses to admit a substratum in the apple on my desk, hanging the apple's entire existence on its properties. But for there to be something really there on my desk, the properties at least should really be present. Yet the empiricist denies the properties any categorical basis, hanging *their* existence on causal dispositions or powers. But it is odd to think of dispositions as substantially present – they are after all mere potentialities. The apple on my desk, however, seems to be more than a cloud of potentialities – something is actually there! Furthermore, the causal dispositions are just dispositions to entertain regularities. But regularities between *what* and *what?* Given that we still don't have substantial objects and properties in the empiricist's world, it's unclear

what the relata of these regularity relations are supposed to be. In other words, the empiricist's world lacks any substantial grounding in something with real presence.

We may think of this as a problem of *circularity*. The empiricist metaphysician promises (perceptual) insight into the nature of objects in terms of properties, then insight into the nature of properties in terms of causation, then insight into the nature of causation in terms of regularity, but finally offering us no independent insight into the relata of this regularity. If we think of the regularities as holding between objects, or objects' properties, then we're led back to the notions we are trying to understand. When you fill in the details, then, it turns out that for the empiricist an object is just a cluster of potential regularities between objects acquiring or losing properties. Embarrassingly, this appeals to the notion of an object to elucidate what an object is!

For the empiricist vision to become substantive and informative, the tight circle Hume constructed requires some entry point from the outside. The nature of either objects or properties or causation must be construed as outrunning the perceptible phenomena. One could, for example, revert to the view that properties are not causal dispositions, but the categorical bases of such dispositions. One could then understand objects as clusters of categorical bases of causal funds, and causation as regularity among events in which these clusters acquire a new member or lose an old one. Unlike the consistent, all-encompassing empiricism of Hume, this kind of limited empiricism is a stable position.

The problem with it, however, is that once the principled interdiction on para-perceptual insight into "noumenal" nature has been lifted, it is unclear why we should restrict this kind of insight to only one ontological category. After all, if there is something deeply illegitimate about this, as Hume seemed to think, then we should not do it at all; conversely, if we allow ourselves to do it somewhere, why not everywhere? It would seem somewhat arbitrary to insist that we must restrict our conception of objects and causation to the perceptible phenomena, even though we can appeal to an imperceptible noumenal nature

in the case of properties. We might as well, in other words, conceive of objects, properties and causation alike in terms of that deep noumenal nature.

## 6. Rationalism and the opacity of nature

At first glance, the rationalist's world is much more reassuringly stable. The fundamental pattern of objects with properties causally interacting is construed as follows: there are substrata that carry categorical properties, such that sometimes one substratum's acquisition or loss of a categorical property produces the acquisition or loss of categorical properties in another substratum. For example, when fire causes smoke, what we see is that yellowish-orange flame-shapes are closely followed by grayish-white fume-shapes. But that is just the *phenomenal* transaction. The *noumenal* transaction is this: a fiery substratum, in virtue of its categorical properties, produces ("secretly") a smoky substratum with its own categorical properties. *This* is what really takes place in the world.

This rationalist picture, unlike the Humean one, does not lead us in a futile circle. On the face of it, each component – the objects, the properties, the causal relation – is grasped independently of the others, in terms of its deep essence.

On closer inspection, however, the insight we have into the nature of each component is disconcertingly thin. After all, we don't really know what a substratum is, beyond the fact that it's a *je-ne-sais-quoi* bearing properties; we do not know what a categorical property is, beyond that it's a *je-ne-sais-quoi* supporting causal dispositions; and we do not know what production is beyond that it secretly underlies regularity relations. To speak of these metaphysical elements, we use such comfortingly familiar words as 'substratum', 'basis' and 'producing'. But notwithstanding the fact that in everyday life these words can be used to speak of things into which we have some perceptual insight, here we use them to signal essentially unknowable things. What these words mask is the dispiriting fact that the rationalist attempts to explain the causal transaction between fire and smoke in terms of three separate types of *je-ne-sais-quoi*.

Flowery, suggestive language aside, what the rationalist *really* tells us is just that an object is a *je-ne-sais-quoi*<sub>1</sub>, a property is a *je-ne-sais-quoi*<sub>2</sub>, and causation a *je-ne-sais-quoi*<sub>3</sub>. When fire causes smoke, then, what really happens, according to the rationalist, is this: a JNSQ<sub>1</sub> that supports some JNSQ<sub>2</sub> bears a JNSQ<sub>3</sub> relation to another JNSQ<sub>1</sub> with its own JNSQ<sub>2</sub>. Hardly an illuminating account of the fundamental patterns of our world! We might as well be told that a blah with bleh bears bluh to another blah with its own bleh. (On second thought, that *is* what we are told here!)

In truth, then, the rationalist offers us no genuine insight into the nature of objects, properties and causation. She leaves the fundamental patterns surrounding us completely opaque. The empiricist may have led us in a circle, but at least she made a good-faith effort to provide illumination. The rationalist just uses familiar words to mask the essential opacity of nature as she conceives of it.

## 7. Enter introverted empiricism

Traditional metaphysics has spontaneously organized around three domains of existence: physical, mental, and abstract. Here too, one suspects an underlying epistemological trichotomy: sense perception, introspection, and reason. In the first instance, physical entities are paradigmatically accessed through sense perception, mental entities are paradigmatically accessed through introspection, and abstract entities are paradigmatically accessed through reason.

As we have seen, the central debates on the nature of objects, properties, and causation are also epistemologically grounded, contrasting a Humean empiricist approach that attempts to offer a perception-based insight into their nature and a rationalist approach that attempts to offer a reason-based insight. It would seem, though, that there is a third option we have not considered, namely, appealing primarily to introspection to anchor insight into objects, properties, and causation.

The nature of introspection itself is a controversial matter, of course, and here I will make one important assumption about it. This is the assumption that

introspection is crucially like perception in that it involves a kind of direct encounter with its objects (though it may also be crucially unlike perception, insofar as its operation is not grounded in a dedicated organ). Seeing a dog is different from thinking about a dog in that it presents the dog in a distinctively direct way, “*in persona*” as Husserl put it. Likewise, introspecting a headache is very different from just thinking about a headache, and presents the headache *in persona* as well. In that respect, introspecting a headache is a kind of inner-perceiving of it.

This feature of introspection raises the specter of a third approach to our insight into the nature of objects, properties, and causation, an approach that is empiricist without being Humean. The approach would anchor our insight in introspection, construed as a kind of inner encounter with *mental* objects, *mental* properties, and *mental* causation. Insofar as the source of insight is construed as perception-like, the view is empiricist. But since the relevant perception is not sense perception, this is not Humean empiricism. To reflect these two facets of the view, I propose to call it *introverted empiricism*. I now turn to discuss introspective empiricism’s take on the nature of objects, properties, and causation—in reverse order.

## 8. Introverted empiricism and the nature of causation

Introverted empiricism is applied to causation by Franz Brentano. There is more to causation than regularity, insists Brentano, and while in sense perception we cannot perceive the extra factor (the “secret connexion”), we can do so in introspection:

... so far as cause and effect are concerned, we should consider those cases in which causation is intuitively [that is: directly] apprehended. We find these only in the domain of inner experience. An example is the case where we derive a judgment from certain premises. We note that the conclusion is made self-evident to us, not only after the premises, but also from the premises. Insofar as we think the premises, we experience ourselves as active; insofar as we apprehend the conclusion from the premises, we experience ourselves as passive and acted upon.... (*Kategorienlehre*, pp. 55-6).

Brentano seems to claim that when one thinks that  $p$  and that  $p \rightarrow q$ , and is thereby led to judge that  $q$ , one experiences the thought that  $q$  not only as occurring after, but also as *produced by*, the thoughts preceding it. One experiences not only the *post hoc* but also the *propter hoc*.

For Brentano, this means that our concept of causation is grounded in inner perception of experienced production. Only once we have thus acquired the concept of causation do we apply it to transactions outside the mental domain:

On the basis of this inner experience of being caused, we assume by analogy that such causation also occurs in many cases of regular succession where there is no such immediate manifestation of causation. (*Kategorienlehre*, p. 56).

We first notice that, in the mental domain, a stubborn symptom of one type of thought producing another is the “regular succession” of the producing and produced thoughts. Sense perception reveals the same pattern in external events – for example, when smoke follows fire with “regular succession” – but without also revealing the production of one event by another. Nonetheless, analogical inference leads us to surmise that regular succession must be the symptom of the same kind of intimate transaction in the external world. This is how we come to think of external physical events as genuinely producing one another, even though we have never witnessed physical production.

Brentano’s approach here is a paradigm of introverted empiricism, applied to causation. However, the central claim that we experience our thinking the conclusion as *produced by* our thinking the premises is far from introspectively obvious. Suppose you need something from the store, which you know to close at 8, and then discover that it is already 8:20. The thoughts “The store closes at 8” and “It’s already 8:20” occur to you, and soon thereafter you experience the thought “The store is already closed.” Can you really experience the *causing* of the latter thought by the former ones? My own introspective impression is that I do not.

Brentano might respond that there are certainly cases in which we experience some of our thoughts as occurring *because of* other thoughts. Compare and contrast the following cases. In a first case, you try to remember

the name of the capital of Liberia, which frustratingly lingers on the tip of your tongue, before suddenly popping up in your mind. You then experience the thought “The capital of Liberia is Monrovia” but without experiencing the causal processes that so to speak spit up this thought. In a second case, you encounter in a museum a captivating painting by a painter you have never heard of, say Felix Nussbaum, and, wanting to know more about him, you decide you will google him when you return home. Here you experience the thought “I will google Felix Nussbaum at home” in close succession after “I don’t know this Felix Nussbaum” and “I’d like to know more about him”, but you also experience the first thought as occurring *because* of the other two. Likewise, if you have to decide who will win a football match between Spain and Germany, you entertain a number of relevant considerations – who’s injured and who’s playing, what tactics each team is likely to employ, and so on – and make a judgment, say that Spain will win. Here the thought “Spain will beat Germany” is experienced as occurring not only *after* but also *because* of the thoughts about the players and tactics. The point is that while some thoughts are experienced as merely *post hoc*, others are experienced as *propter hoc*. A Brentanian could hold that we acquire the concept of *because-ness* through introspective encounter with mental transactions of this sort.

This Brentanian response, however, leaves unclear how we obtain any insight into the secret connexion between cause and effect, the cause’s *producing* of the effect. Even when we experience the thought “Spain will beat Germany” as occurring *because* of the prior thoughts about players and tactics, it is hard to point at an experiential element to do with the prior thoughts *producing* the later one. Relatedly, what is missing in the case of the thought about Liberia’s capital is first of all an experience of the cause; the contrast we’re looking for, however, is not between a case where we experience both the effect and the cause and a case where we experience only the effect, but between a case where we experience nothing but the cause and the effect and a case where we experience also the producing of the effect by the cause.

Perhaps it would be more plausible to hold that production is introspectively manifest in a special kind of causation found only in the mental domain, what is sometimes called “agent-causation” (as opposed to “event-

causation"). In most causal exchanges, both the cause and the effect are events. But in the mental domain, some philosophers have claimed, there are causal exchanges in which the effect is an event but the cause is a person. When I decide to imagine a two-headed camel and then do so, I do not experience *the decision* as causing the image. I experience *myself* as causing it. Perhaps the experience is of me-qua-decider causing the image, or of me causing the image *in virtue of* deciding to do so; still, it is me who is experienced as the cause. Indeed, I experience myself specifically as producer of the image, and experience the producing of the image by myself.

It is hard to assess the plausibility of this claim, but it does seem more phenomenologically plausible than the claim that we experience some of our thoughts as producing others. An introverted empiricism could claim that our original insight into the causal nexus is obtained through introspective encounter with this kind of "agent-production". Having acquired the concept of causation by such direct encounter with something that goes beyond regularity, we then apply the concept in cases in which all we directly witness are regularities.

## 9. Introverted empiricism and the nature of properties

If causation is construed as production, then dispositional essentialism about properties becomes more tenable, casting properties as funds of dispositions to produce some effects. But introverted empiricism gives us the tools to defend a more robust account of properties as involving categorical bases.

The observation that empirical science only tells us about physical properties' causal relations (actual and potential) to each other, remaining silent on their (categorical) intrinsic natures, was made already by Bertrand Russell. Russell went on to speculate that the intrinsic, categorical *je-ne-sais-quoi* of physical properties (as described by physics) was cut of the same cloth as that of mental properties. Accordingly, he suggested grounding both physical and mental properties in more basic properties that are in themselves neither mental nor physical.

Russell's speculation presupposes that the intrinsic dimension of a property is equally opaque in the mental and physical cases. Modern variations on the Russellian speculation have instead underscored an asymmetry between the two: in the case of mental properties, we sometimes encounter in introspection not only their (actual and potential) causal relations, but also their intrinsic natures. We know that pain attracts attention, functions as alarm for tissue damage, and so on; but in addition, we introspectively encounter that distinctively hurtful intrinsic character of pain. So while the intrinsic nature of physical properties remains an opaque *je-ne-sais-quoi*, that of (some) mental properties can be directly grasped. This has led some to speculate that the categorical basis of physical properties as empirically described just is the intrinsic experiential feel we know from introspection. David Chalmers writes:

Russell pointed out that physics characterizes physical entities and properties by their relations to one another and to us. For example, ... a property such as mass is characterized by an associated dispositional role, such as the tendency to resist acceleration. At the same time, physics says nothing about the intrinsic nature of these entities and properties. Where we have relations and dispositions, we expect some underlying intrinsic properties that ground the dispositions, characterizing the entities that stand in these relations.... [On the other hand, some mental] properties seem to be intrinsic properties that are hard to fit in with the structural/dynamic character of physical theory; and arguably, they are the only intrinsic properties that we have *direct knowledge of*. Russell's insight was that ... the intrinsic properties of the physical world are themselves [mental] properties. Or perhaps the intrinsic properties of the physical world are not [mental] properties, but nevertheless constitute [mental] properties: that is, perhaps they are [proto-mental] properties.... Physics as we know it emerges from the relations between these entities, whereas consciousness as we know it emerges from their intrinsic nature. (Chalmers, "Consciousness and Its Place in Nature", p. 133; my italics)

In this passage, Chalmers considers two possible speculations. The first is that the categorical bases of causal dispositions are intrinsic mental properties. The second is that they are *proto-mental*, that is, intrinsic properties that in themselves are neither mental nor physical but underlie both types of property. The first speculation is problematic insofar as it leads to a form of panpsychism: if there is a mental property underlying every physical property, then even rocks and molecules have mental properties. The second speculation does not have this implication, but is problematic insofar as it leaves us with no insight into the

nature of the categorical basis, since we have no idea what these proto-mental properties are like.

There is a more cautious and less speculative way to develop this line of thought, however. The notion of categorical basis is originally formed on the basis of introspection of mental properties. It is then applied to physical properties accessed through sense perception (directly or indirectly). As noted, we know that pain plays an important functional role, but we can also directly introspect its intrinsic feel. Furthermore, this intrinsic feel seems to *ground* and *explain* the functional role: pain attracts attention to tissue damage because it feels so bad, in the same sense a bomb is explosive because it contains potassium nitrate. Having witnessed the categorical basis that grounds the causal disposition in the mental case, we can infer by analogy the existence of some categorical basis in the case of physical properties, though it need not be the same, *mental* categorical basis.

Importantly, it is no part of this suggestion that we can know how strong the analogy is between mental and physical categorical bases. On the one end of the spectrum, the analogy is perfect, so that the categorical bases of physical properties are themselves intrinsic feels (as per panpsychism). On the other end of the spectrum, the analogy may be very weak, so that the categorical bases of physical properties have nothing to do with feels, and the only similarity pertains to their function as grounding bases. In that scenario, introspection gives us insight into the categorical basis of only some properties, though it gives us reason to believe that all dispositional properties are ultimately categorically grounded. This is still better than what Humean empiricism and rationalism offered us: the former promised access only to causal dispositions, the latter posited unknowable categorical bases but gave us insight into the nature of not a single one among them; the whole categorical domain remained entirely opaque. In that respect, introverted empiricism represents progress over both views.

## 10. Introverted empiricism and the nature of objects

If properties are construed as categorical bases of production funds, then at the very least, objects can be seen as bundles of such categorical properties. This gives them more substance than they seem to have when construed as baseless bundles of regularity funds. In fact, however, introverted empiricism may open the door to a more robust metaphysics of objects as involving substrata that support the relevant bundles.

When you perceive a square red table, all you seem to be aware of are the table's various properties: its being square, its being red, and so on. But when you have a headache, arguably you are also aware of *yourself* as the thing that is *having* the headache. The full experience is not just "pain is happening"; it is "I am in pain". It would be quite pathological to experience your pains as floating qualia occurring in some impersonal "inner space" (whatever that might mean). Normally, we experience our headaches not as impersonal hurts; we experience them as hurting *us*. That is, we experience the headache as a property and ourselves as the substratum that has that property.

Arguably, this is why when we take distance from a headache, stop "identifying with it" and start "observing" it as an object among others, it slowly loses its "edge." The standard experience of pain is one in which the pain is fused with the self pained by it. Indeed, standard pain could be described without excessive injustice as an experience of oneself *modified* in a distinctive way, the painful way. To that extent, the experience of pain not only reveals the property of being in pain but also involves a direct awareness of the bearer of that property, that which *has* the pain.

Hume famously presented a completely different take on introspective awareness of the self:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception. (*A Treatise of Human Nature*, I.iv.6).

This passage runs together two different claims: the first is that we are never aware of ourselves without also being aware of some experience or another; the second, that the experience in question exhausts the relevant awareness. The

first claim is relatively moderate and is fully consistent with the thought that we experience our headaches as our own. The second is much more radical and rules out any involvement of the self in what one is aware of.

How plausible is Hume's stronger claim? It depends in part on what Hume expected to find when looking for "the self". Suppose you look at the rainbow while eating a cheese-and-tomato sandwich, and this puts you in an introspective mood in which you pay closer attention to your own stream of consciousness. You notice the different color qualia in your experience, the cheese and tomato qualia, the quality of pleasant peacefulness enveloping you, and so on. If you look in addition for a "self quale" on a par with these other qualia, you are bound to be disappointed. The self, I propose, shows itself in experience in a completely different way – not a separate quale of its own but as a structural dimension of all qualia. When you have a visual experience of the yellow part of the rainbow, there is a yellowish way it is like for you to have your experience. When you have a gustatory experience of the cheese in your sandwich, there is a cheesy way it is like for you to have your experience. These experiences are partly similar and partly dissimilar. They are dissimilar in that the visual yellowish quality is very different from the gustatory cheesy quality. But they are similar in that in both cases it is *for you* that there is something it is like to have the experience. We might say that a "yellowish way it is like for me" has two components, the yellowish component and the for-me component. The former component varies from one experience to the next, but the latter component is invariant across all experience. Every experience we have is experienced as ours in the sense that there is something it is like *for us* to have it. It is here that the self shows up in experience – not as a detachable stand-alone quale like the yellowish and cheesy qualia, but as a structural dimension of every possible quale.

The absolute universality of the for-me-ness of experience explains, I suspect, its relative elusiveness. Typically, we notice aspects of experience by introspectively contrasting cases in which they are present and cases in which they're not. When such a contrast is unavailable, it is very hard to notice even the most intimate aspects of experience. Consider the incessant hum of the refrigerator in the background as you are engrossed in reading. Typically, you

do not notice the hum until it stops – it is the stopping that creates the kind of contrast between presence and absence that makes you aware both that you no longer experience the hum and that you had been experiencing it for some time. Now imagine if you will a world – call it ‘Fridge World’ – just like ours but for this minor detail: there is a background humming sound accompanying every person from birth to death. In all likelihood, in Fridge World we would be totally unaware of this aspect of our auditory phenomenology. Yet it would be a very real aspect: if the hum were to suddenly stop, we *would* notice a change in our phenomenology. Something like this may well be the case with our experience of ourselves as subjects of experience. Brentano certainly thought so:

Noticing ... presupposes, however, that we encounter in our consciousness privative or positive contrasts to what we are meant to notice. But this cannot a priori be expected to happen in each and every case. What should prevent there being a certain element which exists generally in the phenomena of our consciousness in the sense that each of them participates in it[?] ... I have already mentioned that we have reasons to believe that there actually is such an unnoticeable part in us. We do not understand ourselves as [given] in an abstract concept, but as [given] in a concrete, individual intuition, and yet we are incapable of giving an account of the individuating factor. (*Deskriptive Psychologie*, pp. 61-2; cf. *Kategorienlehre*, p. 160)

Since the self is a component of every possible experience, no contrast is available between cases in which it is present and cases in which it is absent. In consequence, if like Hume you look for it the way you look for yellowish qualia and cheesy qualia, you will not find it. Nonetheless we experience it every moment of our waking life. The introverted empiricist’s hypothesis is that it is in virtue of this inner experience of one substratum which *has* (experiential) properties that we conceptualize objects as more than just bundles of properties. In other words, each of us encounters exactly one object whose noumenal nature she experiences directly – namely, herself. But this gives us insight into the noumenal nature of an object as such.

## Conclusion

We started with a metaphysical question about the nature of objects, properties, and causation – arguably the most fundamental patterns in the world as we

experience it. We saw that debates about the nature of these entities tend to organize around two poles, a Humean empiricist pole and a rationalist pole. The former appeals exclusively to what is perceptually accessible to us but seems to leave something unintelligible in objects, properties, and causation. The latter renders objects, properties, and causation more intelligible but leaves their ultimate nature entirely opaque. I have suggested that a relatively neglected approach to the metaphysics of objects, properties, and causation might be able to combine the advantages of both views, anchoring our grasp of the ultimate nature of objects, properties, and causation in direct, quasi-perceptual introspective encounter with selves, intrinsic feels, and agent-production. This approach thus holds the promise of casting objects, properties, and causation as fully intelligible but also as having intrinsic natures which are at least *sometimes* directly manifest to us. An interesting consequence of introverted empiricism is that there is a kind of intimate self-understanding that underlies our understanding of the ultimate character of reality.

Introverted empiricism faces some challenges. One type of challenge is to make the case that in introspection we really do encounter the self as substratum, intrinsic feels as categorical bases, and/or agent-production. Above, I have sketched some preliminary considerations in support of these claims, but each would require a much more sustained defense to be truly compelling. A second type of challenge is to show how the extension from the mental to the non-mental domain actually works: how and to what extent we can apply our introspectively based conception of the nature of mental objects, properties and causation to the external world. Thirdly, the introverted empiricist should explain what it is about introspective encounter with objects, properties, and causation that allows it to *reveal* their ultimate nature.

These challenges are real and daunting. In this paper, I have attempted only to articulate introverted empiricism and to motivate the notion that it merits further pursuit. Addressing the challenges above in a more sustained way would be one way of developing the program of introverted empiricism.<sup>3</sup>

## References

- Armstrong, D.M. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Brentano, F.C. 1933. *Kategorienlehre*. Ed. A. Kastil. Hamburg: Felix Meiner.
- Brentano, F.C. 1982. *Deskriptive Psychologie*. Hamburg: Felix Meiner.
- Chalmers, D.J. 2002. "Consciousness and Its Place in Nature." In his *The Character of Consciousness*. Oxford and New York: Oxford University Press, 2010.
- Hume, D. 1739-40. *A Treatise of Human Nature*. London: John Noone.
- Hume, D. 1748. *An Enquiry Concerning Human Understanding*. London: A. Millar.
- Kant, I. 1781/1787. *Kritik der reinen Vernunft*. Riga: J.F. Hartknoch.
- Nagton, R. 1998. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Oxford University Press.
- Popper, K.R. 1990. *A World of Propensities*. Bristol: Thoemes.
- Russell, B. 1927. *The Analysis of Matter*. London: Kegan Paul.

---

<sup>1</sup> One view is that a full description of the universe would require mentioning other types of entity, for example events, states of affairs, or spacetime. Another view is that these entities themselves are but different combinations of objects, properties, and causal relations, thus admitting of ontological assay in terms of them. This disagreement does not concern us here.

<sup>2</sup> Note a structural asymmetry between RT and PT: for RT, type-causation is primary and token-causation is derivative; for PT, token-causation is primary and type-causation derivative.

<sup>3</sup> This work was supported by grant ANR-17-EURE-0017. For comments on a previous draft, I am grateful to Arnaud Dewalque, and for useful discussions, to Tim Bayne and Lizzie Kriegel.