Essentialist modal rationalism

Philip Goff[1]

Abstract

It used to be thought that rational coherence and metaphysical possibility went hand in hand. Kripke and Putnam put a spanner in the works by proposing examples of propositions which seem to violate this principle. I will propose a nuanced form of modal rationalism consistent with the Kripke/Putnam cases. The rough idea is that rational coherence entails possibility when you grasp the essential nature of what you're conceiving of.

Keywords Conceivability · Possibility · Modal rationalism · Physicalism · Consciousness · Zombies · Hard problem · Two-dimensional semantics

In my recent book *Consciousness and Fundamental Reality*, I proposed a principle linking rational coherence and metaphysical possibility, as part of an argument against physicalism. Although it was not the focus of concern in this book, I had hoped that that principle might undergird a generalised account of our knowledge of modality. I have subsequently realised, however, that that principle has limited application, in a way that conflicts with these broader ambitions. In this paper I will outline these limitations and propose ways of overcoming them. The result, I hope, is the bare bones of an account of how our knowledge of metaphysical modality is grounded in our capacity to discern whether or not a proposition is rationally coherent.

In Sect. 1, I will give a brief overview of the view I defended in *Consciousness and Fundamental Reality*. In Sect. 2 I will explain its limitations and propose a way of resolving them.

✉ Philip Goff
philip.a.goff@durham.ac.uk

1 University of Durham, Durham, UK

1 Introduction

It used to be thought that rational coherence and metaphysical possibility went hand in hand, something we can capture with the following principle:

> *Simple Modal Rationalism*—For any proposition P, P is rationally coherent (where a proposition is rationally coherent just in case it can't be ruled out a priori) iff there is a metaphysically possible world at which P is true.

Kripke (1980) and Putnam (1973) put a spanner in the works by proposing examples of propositions which seem to violate this principle, such as <water is XYZ>.[1] We cannot know a priori that water is not XYZ, and yet there is no genuine possibility corresponding to this state of affairs. In *Consciousness and Fundamental Reality*, I proposed a nuanced form of modal rationalism consistent with the Kripke–Putnam counterexamples to Simple Modal Rationalism. The rough idea is that rational coherence entails possibility when you grasp the essential nature of what you're conceiving of. We can call this general approach 'Essentialist Modal Rationalism'.

Water is essentially H_2O and it is thus not possible for it to be XYZ. Why then is it coherent to suppose that water is XYZ? The reason is that the concept <water> does not reveal the essential nature of water. When conceiving of water as 'water' we don't conceive of it in terms of its essential nature (being composed of H_2O molecules) and thus it is coherent to suppose that it has some other essential nature (being composed of XYZ molecules). According to the modal rationalism I want to defend, coherence and possibility go together when the concepts being employed yield complete understanding of what is being conceived of (from hereon, by 'possibility' I mean metaphysical possibility).

In *Consciousness and Fundamental Reality*, I expressed this in terms of the distinction between *transparent* and *opaque* concepts, defined as follows:

- *Transparent Concept (1st attempt)*—A concept C referring to entity E is transparent just in case C reveals the nature of E (i.e., what it is for E to be part of reality is a priori accessible for someone possessing C, in virtue of possessing C), for example, <sphericity> and <party>.
- *Opaque Concepts*—A concept C referring to entity E is opaque just in case C reveals little or nothing about the nature of E, for example, <water> and <gold> (Goff 2017: 74).

For sphericity to be part of reality (by being instantiated[2]) is for there to be something with all points on its surface equidistant from its centre; for someone possessing the concept of sphericity, this can be known a priori. For water to be part of

---

[1] Kripke (1980), Putnam (1973).

[2] The form of reality will depend on the kind of entity. For an individual to be part of reality is for it to exist; for a property to be part of reality is for it to be instantiated (Platonists might want to distinguish a property's existing from its having concrete reality); for an event to be part of reality is for it to take place.

reality is for there to be something composed of H_2O molecules; in contrast to the case of <sphericity>, this cannot be known a priori in virtue of possessing <water>.

We are now in a position to give a first attempt at defining a special notion of coherence:

Transparent Coherence (1st attempt)—For any proposition P, P is transparently coherent just in case:

(i) P contains only transparent concepts, and
(ii) P is rationally coherent, i.e. the truth of P cannot be ruled out a priori.

Building on this, I defended the following principle linking coherence to possibility:

Transparency Coherence Principle (1st attempt)—For any proposition P, if P is transparently coherent, then P is possibly true.

And on the basis of this principle, I built a case against physicalism, roughly:

The transparency argument against physicalism

1. *Transparency Coherence Principle*—For any proposition P, if P is transparently coherent, then there is a metaphysically possible world at which P is true.
2. The proposition <there are zombies> is transparently coherent.
3. If there is a metaphysically possible world at which <there are zombies> is true, then physicalism is false.
4. Therefore, physicalism is false.

The defence of premise 2 involved arguing that the concepts involved in <there are zombies>, namely physical and experiential concepts, are transparent.

2 Essentialist Modal Rationalism

2.1 Accommodating brute one offs

The Transparency Coherence Principle outlined above is a one-way conditional, and as such it does not rule out that there are propositions that are possibly true but not transparently coherent (in the sense defined above). If there are such propositions, and we know that they are possibly true, then we cannot account for our knowledge of their possible truth in terms of our knowledge of their transparent coherence (because *ex hypothesi* they aren't transparently coherent, at least not in the above sense). And hence, if we want a general account of our modal knowledge in terms of our knowledge of transparent coherence, then we need the Transparency Coherence Principle in a biconditional form:

> *Biconditional Transparency Coherence Principle*—For any proposition P, P is transparently coherent iff P is possibly true.[3]

Unfortunately, so long as we are understanding rational coherence as defined in Sect. 1, there are counterexamples to the above principle. To explain why, we need to introduce some more terminology. Let us say that an entity has a 'defining nature' just in case it is possible to understand what it is for that specific entity to be part of reality. Kinds (e.g. water) and properties (e.g. sphericity) seem to have defining natures; we can grasp what it is for these things to be instantiated. But it is not clear that particular individuals, such as a particular electron E, have defining natures. E falls under a kind—the kind electron—and that kind has a defining nature. But there is nothing to be grasped about what it is for *E*, as opposed to some other electron, to exist. In this sense, particular individuals are 'brute one offs.' (Or at least that is how it seems. Perhaps particular individuals do have defining natures but grasping them is beyond our ken. Perhaps God knows each thing in terms of its defining nature).

For any entity that lacks a defining nature, the notion of a transparent concept of that entity gets no purchase. If there is nothing to grasp about what its reality consists in, then there can't be a concept that reveals what its reality consists in. This leads to difficulties for the Biconditional Transparency Coherence Principle. Let us assume that a human individual lacks a defining nature.[4] Now consider the proposition <Lewis Carroll exists but Karl Marx does not>. This proposition is both coherent and possibly true. But it is not *transparently* coherent, as it contains opaque concepts. Nor is it possible to replace the non-transparent concepts involved with co-referring transparent concepts, given that these concepts denote entities that lack defining natures. Thus, we have a genuine possibility—the possibility of Lewis Carroll existing without Karl Marx—which does not correspond to a transparently coherent proposition.

In *Consciousness and Fundamental Reality*, I tried to circumnavigate this problem simply by focusing on those propositions which *are* rationally coherent in the above sense, and defining the link between coherence and possibility in terms of those propositions. The problem with this strategy, as stated above, is that it prevents us giving a general account of modal knowledge in terms of knowledge of transparent coherence.

I now think there is a fairly straightforward way of extending the formal definition of transparent coherence to avoid this difficulty. The informal idea of transparent coherence is that it's what a proposition has when it *reveals to the thinker the full*

---

[3] If we specifically want to account for our modal *knowledge*, then strictly speaking all we need is the following principle: For any proposition P, P is transparently coherent iff P is known to be possibly true. However, I can't see that this further qualification will make much difference to the discussion.

[4] In fact, it is somewhat plausible that a human person has a defining nature in terms of the sperm and egg from which it was formed. In this case, that sperm and that egg will either be brute one offs or will also have defining natures in terms of their origins and their DNA. The entities that define those origins will either be brute one offs or have defining natures given in terms of their origins. It seems plausible to me that as we trace this back at some point we'll get back to brute one offs. In any case, we could simply change the example to avoid this worry (if there are no examples of brute one offs, then we can't change the example, but in that case the problem goes away).

reality of what she is conceiving of (when employing the proposition). Now if you don't know whether or not 'Lewis Carroll' and 'Karl Marx' co-refer, then that's a sense in which you don't fully grasp the state of affairs you're conceiving of when you entertain the proposition <Lewis Carroll exists but Karl Marx does not>. For all you know, you might be conceiving of one person existing without himself, but equally you might be conceiving of one person existing without some other person. In so far as this proposition does not reveal to the thinker that Lewis Carroll is distinct from Karl Marx, it fails fully to reveal to the thinker what state of affairs she's conceiving of.

The obvious way to fill in this knowledge gap is to conjoin the proposition with the true information about whether or not Lewis Carroll and Karl Marx are identical. Doing this would remove the ignorance discussed in the previous paragraph whilst leaving us with a coherent proposition:

≪Lewis Carroll exists without Karl Marx> & <Lewis Carroll is not identical with Karl Marx≫

Of course, if it had turned out (epistemically) that Lewis Carroll *was* identical with Karl Marx, then adding that information would have produced an incoherent proposition:

≪Lewis Carroll exists without Karl Marx> & <Lewis Carroll is identical with Karl Marx ≫[5]

Whether or not instances of the above form of proposition are possibly true depends on whether or not the entities referred to are identical, and adding this information gives us an incoherence in cases where the truth of the original proposition was impossible and a coherence in cases where the truth of the original proposition was possible. In other words, we've brought coherence and possibility back together. This suggests a way to extend the definition of transparent coherence:

Transparent coherence (refined definition)

- Proposition P is transparently coherent just in case:
- (i) P is coherent,
- (ii) In so far as P contains concepts referring to entities with defining natures, those concepts are transparent,
- (iii) In so far as P contains concepts referring to brute one offs B_1, B_2… B_n: P* is coherent, where P* is the proposition formed from conjoining P with true information concerning all identities and non-identities between the entities involved in B_1, B_2… B_n.

With this extended definition of rational coherence, the problem is avoided. The proposition <Lewis Carroll exists without Karl Marx> is transparently coherent and corresponds to a genuine possibility; the proposition <Lewis Carroll exists without

---

[5] I am assuming here that the necessity of identity is a priori.

Charles Dodgson> is not transparently coherent and does not correspond to a genuine possibility. Exactly the results we wanted.

One might wonder why the *modal* constraints on brute one offs are confined to facts concerning identity/non-identity. In fact, there may well be further modal constraints given by the essential kinds brute one offs falls under. If E is essentially a member of the kind *electron*, then E will be an electron in any possible world in which it exists. For E to lack a defining nature is for E's *specific* identity—its identity over and above the kinds it falls under—to be brute: there is nothing to be grasped about what it is for E to exist as opposed to some other electron. It follows that the only constraints on E's identity, beyond being an electron, are those given by logic and the necessity of identity. Some may want to analyse brute identity in terms of the possession of an individuating property—a haecceity—or a substratum. Alternately one may simply accept that there are brute facts about identity and non-identity lacking any deeper underpinning. I need not decide between these options for my purposes here.

2.2 Accommodating Ostrich nominalism

This modified definition of transparent coherence improves on the original by virtue of accommodating a new kind of entity: brute one offs. However, it still retains an exclusive focus on *entities*, which leads to a further difficulty for the Essentialist Modal Rationalism we are building upon it. Consider the view David Armstrong mockingly dubbed 'Ostrich nominalism', according to which there are no properties but there are irreducible facts about how entities are. For the Ostrich nominalist, there is no such thing as redness but nonetheless roses are red, and there is no such thing as negative charge but nonetheless electrons are negatively charged. We might put the view by saying that whilst there are no such things as *properties*, there are such things as *propertied objects*: red roses, charged electrons, etc.

I have defined transparent concepts as a subset of entity-referring concepts: they are those entity-referring concepts that reveal the nature of the entities they refer to. But if there are no properties, then there are no entity-referring concepts associated with predicates, and hence there are no transparent concepts associated with predicates (given the above definition of a transparent concept). And, of course, if there are no such things as properties, then we cannot bring in facts about property identity to determine whether a coherent proposition is transparently coherent, as instructed in the third clause of the refined definition of transparent coherence.

Why is this a problem? Assume Ostrich nominalism is true. It would follow that essentialist modal rationalism would be inconsistent with Kripkean exemptions to Simple Modal Rationalism. Consider the proposition: <The material of the golden broach has atomic number 8>. This proposition is clearly coherent. And for the reasons discussed above, if there are no properties associated with its predicates, then we cannot rule out its transparent coherence with reference to the identity of those properties or how they are conceived of. Hence, it seems we must conclude that this

proposition is transparently coherent and, assuming the Biconditional Transparency Coherence Principle, that there is a possible world at which it is true.

For those who accept the Kripkean exemptions to Simple Modal Rationalism, this is the wrong result. Strictly speaking, the transparency argument against physicalism would go through—as propositions concerning zombies would end up being transparently coherent for the same reason—but the victory would be won too easily. We cannot rule out psycho-physical empirical identities by ruling out all of the standardly accepted instances of empirical identities, such as <water is H_2O> and <Gold is the element with atomic number 79>.

Of course, we could just assert that essentialist modal rationalism is committed to the existence of properties, but this would give the view limited application. A better way forward is to find a more nuanced definition of transparent concept such that predicates can express transparent concepts even in the absence of properties. This is what I will try to do in what follows.

I take it that, for the Ostrich nominalist, the predicates of true sentences capture something about reality, just not something that amounts to the presence or absence of entities.[6] Imagine an Ostrich nominalist world W containing only a red, spherical object. If you only know that there is an object in W, then you are partially ignorant about W: you don't know that the object in W is red and round. Your ignorance does not consist in failing to know that certain entities exist: there is only one entity and you know that it exists. Nonetheless, your ignorance concerns the metaphysical reality of W.

In other words, for both the Ostrich nominalist and the property realist, there is something that is known about reality in knowing that a given predicate, say, charge, truly applies to something in the world. For the property realist, knowing what is ascribed by 'charge' will be a matter of knowing the essential nature of charge; for the Ostrich nominalist, it will be a matter of knowing what it is for something to be charged.

We can perhaps get clearer on this with reference to Theodore Sider's notion of *metaphysical truth conditions.* The metaphysical truths conditions of a sentence are the truth-conditions of S in a perfectly natural language, that is to say a language involving only terms that perfectly carve nature at the joints. Metaphysical truth conditions apply to whole sentences, but we can think of a specific term in terms of its *metaphysical contribution*, that is to say, the contribution it makes to the metaphysical truth conditions of sentences in which it is involved. If 'negatively charged' is a perfectly natural predicate, then its metaphysical contribution can, in part, be given as follows: For any sentence S of the form 'X is charged', S is true iff X is negatively charged.

To know the metaphysical contribution of the predicate 'negatively charged' is know what it is for something to be charged. Distinct predicates can make the same metaphysical contribution; and while some predicates will be a priori associated

---

[6] This needs to be qualified. Even for an austere nominalist, the successful application of some predicates will be a matter of the presence of entities, e.g. 'is identical with Kripke'. I am here focusing on those predicates that property realists take to correspond to perfectly natural properties.

with their metaphysical contribution, others will not. Contrast the following two sentences:

(A) 'The table is spherical'
(B) 'The table is as David thinks it is', where David actually thinks the table is spherical, and the meaning of the predicate is such that its metaphysical contribution across all possible worlds is determined by how David actually thinks the table is (i.e. it is a non-referential analogue of a rigid designator).

In both cases, the metaphysical contribution of the predicate is the same, but in the former but not the latter case someone who understands the predicate has a priori access to its metaphysical contribution. The person asserting sentence A knows what they are claiming about reality, whereas the person asserting sentence B may not.

In this way, we can give a definition of transparent concept such that a predicate can express a transparent concept even if it doesn't refer to a property:

- *Transparent Concept (refined definition)*—A term T expresses a transparent concept iff T's metaphysical contribution is a priori accessible.[7]

With the refined definitions of transparent coherence and transparent concept, and a biconditional version of the Transparency Coherence Principle, we now have the resources in place to give a robust and highly flexible account of modal knowledge in terms of knowledge of coherence.

I argued in *Consciousness and Fundamental Reality* that an essentialist account of modal rationalism is to be preferred over Chalmers' (2009) two-dimensional account of modal rationalism, as the former avoids the contentious meta-semantic assumptions embedded in the two-dimensional framework. I suppose some philosophers may prefer Chalmers' semantic commitments to the commitments to essence and perfect naturalness employed in my framework.[8] I am persuaded on independent grounds that these latter commitments are indispensable but will not repeat these arguments here. For now, I leave it to others to decide which of these foundational commitments they find more palatable.[9]

---

[7] The corresponding definition of an opaque concept: A terms T expresses a transparent concept iff T's metaphysical contribution is not a priori accessible.

[8] One might wonder why we need to bring in a Siderian commitment to perfect naturalness in order to capture the idea that a predicate 'F' might be a priori associated with what it is for something to be F. The problem is that, so long as we are talking loosely, an austere nominalist must surely allow that there is a property of Fness, and likewise the property realist must allow talk of 'what it is for an object to be F'. What I want talk of 'metaphysical truth-conditions' and 'metaphysical contribution' to hone in on is the most fundamental relationship between representation and reality, relative to various metaphysical theses. It is this that is revealed by a transparent concept, in the case of that concept. If austere nominalism is true, then, at the most basic level, 'X is F' is true because X is F; whereas, if property realism is true, then, at the most basic level, 'X is F' is true because X instantiates Fness. Perhaps there are other ways to capture this kind of priority than investing in perfect naturalness, but it seems to me that something like this commitment is required.

[9] I am grateful to two anonymous reviewers and for advice from the editor of this issue, Antonella Mallozzi.

References

Chalmers, D. J. (2009). The two-dimensional argument against materialism. In B. McLaughlin (Ed.), *Oxford handbook of the philosophy of mind*. Oxford: Oxford University Press.

Goff, P. (2017). *Consciousness and fundamental reality*. New York: Oxford University Press.

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.

Putnam, H. (1973). Meaning and reference. *Journal of Philosophy, 70,* 299–711.