Adam Pautz

Consciousness and Coincidence

Comments on Chalmers

Abstract: In 'The Meta-Problem of Consciousness', David Chalmers briefly raises a problem about how the connection between consciousness and our verbal and other behaviour appears 'lucky'. I raise some questions about Chalmers' formulation of the problem. Then I develop an alternative formulation. Finally, I consider some responses, including illusionism about consciousness.

1. Introduction

Chalmers' *meta-problem about consciousness* concerns the explanation of our *problem responses* to consciousness, where those responses are characterized in physical-functional terms. For instance, why are we disposed to utter sentences like 'consciousness is irreducible' and 'the quality red is distinct from any physical property'?

Given the causal closure of the physical realm, there are bound to be computational processes in the brain that explain such verbal reports, as Chalmers notes (2018, p. 8). What, then, is the *problem*? Chalmers suggests that one problem is that it is *lucky* that these processes are accompanied by consciousness.

I will first raise some questions about *Chalmers' luck problem*. Then I will develop a problem in the vicinity, the *normative harmony problem*. Finally, I will consider some responses, including illusionism about consciousness.

Correspondence:
Email: adam.pautz@gmail.com

2. What is Chalmers' Luck Problem?

Here is Chalmers' most general characterization of his luck problem:

> As long as we have modal independence, so that the meta-problem pro-
> cesses [the computational processes that explain our reports] *could* have
> come apart from consciousness, it can seem lucky that they have
> not…Where realization is concerned, it seems lucky that the meta-
> problem processes are in fact realized by consciousness rather than by
> something else… [I]n the case of consciousness one has a very strong
> sort of modal independence, in that there seems to be a *near-complete*
> structural explanation of the intuitions that could obtain without con-
> sciousness. It is easy to get the sense that what really explains the
> intuitions is the structure of cognitive processes, and the fact that
> consciousness is connected to that structure is something of a fortunate
> and optional extra. (*ibid.*, p. 48)

> Once we can explain our conviction that consciousness exists without
> assuming that consciousness exists, the fact that the conviction is true
> seems somewhat miraculous. (*ibid.*, p. 56)

So, Chalmers says that, given modal independence, the following is
'lucky', even 'miraculous':

> [A] The computational processes that explain our verbal reports
> ('consciousness is irreducible', 'consciousness exists', 'I am con-
> scious of red') are accompanied by consciousness.

He regards this as a significant problem, which may require a radical
solution.[1]

I have two questions. First, what does Chalmers mean by saying that
[A] is *lucky*, and why would this be a *problem*? Does he mean that it
'cries out' for explanation? Would he agree with White (2005, p. 3)
that something is 'lucky' in this sense if it lowers the probability of

---

[1] For instance, Chalmers (2018, p. 48) considers the radical solution that '*only* conscious-
ness could realize the relevant processes, subject to certain constraints'. This solution
apparently rejects modal independence. But I don't see how it could be right —
especially if, like Chalmers, we think conceivability supports possibility. For instance,
given Russellian monism (a view Chalmers has sympathy with), the computational
processes that explain our verbal reports are in fact realized by special proto-
phenomenal properties that (in the right combination) constitute conscious experiences.
But it's conceivable (cannot be conclusively ruled out *a priori*) that they should have
been realized by non-proto-phenomenal properties that are insufficient for conscious
experience, in which case all of us would have been zombies. So isn't it 'lucky' that
they are realized by the special proto-phenomenal properties (Pautz, 2015, Section 4)?

one's previously favoured explanation and raises the probability of some alternative hypothesis?

Second, is Chalmers right in suggesting that, 'as long as we have modal independence', [A] is lucky?

Here is an apparent counter-example. Consider a form of *identity theory* on which conscious experiences are necessarily identical with *carbon-based* biological states P_1, P_2, … which in fact realize the meta-problem computational processes causing our reports (Chalmers, 2018, p. 37). This view upholds modal independence as Chalmers defines it: the computational processes that explain our verbal reports might have 'obtained without consciousness'. For, in an alternative history of the universe, creatures very much like ourselves might have evolved in whom these computational processes are instead realized by silicon-based based states P_1^*, P_2^*, … Now these states are only trivially physically different from our own. Still, on the identity theory, our hypothetical functional-isomorphs would consequently not be conscious, even though, at a certain level of abstraction, their computational processes would be the same as ours and generate the same verbal responses. They merely have 'ersatz-experiences', while we have experiences. (Compare: twater made from XYZ on Twin Earth is not water.)

While the identity theory upholds modal independence, I don't think it must face a deep luck problem. Identity theorists probably should accept 'deflationary pluralism' (Lee, 2019). Our carbon-based states P_1, P_2, … (genuine experiences) are no better (in value, epistemic significance, naturalness) than our hypothetical isomorphs' silicon-based states P_1^*, P_2^*, … ('ersatz-experiences'), because the difference is trivial. For instance, we have reliably true and justified ordinary introspective beliefs about our experiences ('I'm having an orange after-image', etc.); our twins have reliably true and justified ordinary introspective beliefs about their t-experiences. (Compare: we have reliably true beliefs about water; our twins on Twin Earth have reliably true beliefs about twater.) *Given* deflationary pluralism, the fact that the computational processes are accompanied by conscious experiences is not in any sense 'miraculous' or 'lucky'. For, on the present deflationary view, it just amounts to the fact that they are realized by the right kind of carbon-based states — rather than, say, silicon-based states. And this doesn't seem 'miraculous' or 'lucky' in any puzzling sense. In fact, it seems unremarkable. It is unlike a monkey typing 'All the world's a stage'.

So I doubt Chalmers' suggestion that 'as long as we have modal independence' there is a significant problem about miraculous luck or coincidence. This raises a question: if modal independence isn't enough, what assumptions *are* required to generate such a problem?

3. An Alternative Formulation:
The Normative Harmony Problem

I suggest that three assumptions are needed to generate a significant problem about 'luck' or 'coincidence':

> *Distinctness*: Conscious states are distinct from all physical-functional states of persons. On the version I will focus on, they consist in standing in an irreducible relation of conscious representation to 'edenic' qualities that are nowhere instantiated in the world (pp. 25, 34).

> *Essential normative significance*: Conscious states have built-in normative features. For instance, maybe it is in the essence of a state of being ostensibly conscious of a reddish and round thing that it gives you reason to believe (i) that a reddish and round item is present ('dogmatism'), (ii) that you are conscious, (iii) that the quality *red* is simple, etc. Other conscious states have intrinsic value or disvalue; it's in their essences to provide a reason to desire that they continue or stop. Maybe *only* experiences could be the source of such reasons.

> *Causal closure*: The physical realm is causally closed. There is no top-down causation.

To illustrate how these assumptions can generate a 'coincidence' problem, let us begin with a form of *epiphenomenalist dualism* which endorses all three of them. Diagrammatically:
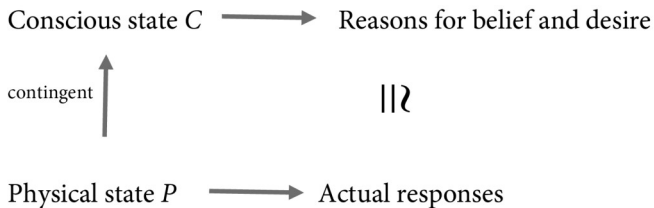

Conscious state *C* ⟶ Reasons for belief and desire

contingent ↑   ‖⁊

Physical state *P* ⟶ Actual responses

Figure 1. Epiphenomenalist dualism

Here, internal physical neurocomputational state P causes certain responses, like 'I'm conscious of red'. The upward arrow indicates *distinctness*, with physical states and conscious experiences being connected by psychophysical laws. These laws might have the form:

> [#] For any x, if x is in physical-functional state P, then x is in the (distinct) state of standing in an irreducible relation of conscious representation to edenic quality $f(P)$,

where f is a systematic function from physical states (e.g. states in a neural similarity-space) onto edenic qualities. The arrow going from conscious experiences to reasons indicates *essential normative significance*. The absence of any downward arrow from our conscious experiences to our responses goes with *causal closure*. For instance, on this view, the non-physical event of feeling of pain in your hand provides you with a reason to withdraw it from the fire, but it is not the *cause* of your withdrawing.

As Chalmers notes, dualism faces a 'coincidence' problem:

> Where psychophysical laws are concerned [e.g. laws of form [#] above], it seems lucky that the laws are as they are. Only this luck ensures that we are not in a zombie world with physical processes and phenomenal intuitions but no consciousness, or in an inverted world where these processes yield pleasure when we feel pain. (2018, p. 48)

But what exactly is the problem here? My formulation of the problem differs from Chalmers' (§1). The problem for epiphenomenalist dualists is not about the 'luckiness' of [A]. Rather, it concerns how to *explain* the following striking regularity, a regularity whose formulation presupposes distinctness and essential normative significance:

> *Normative harmony*: In every case, the psychophysical laws correlate a physical-functional state P with a distinct conscious experience C whose essential normative role in providing reasons is *harmonious* with the causal role of P in generating verbal and other responses.

Call this the *normative harmony problem* (Pautz, 2010, pp. 359, 365).

Here is an illustration. You undergo physical-functional state P that causes you to say '*I'm conscious of the quality red*'. In agreement with normative harmony, a psychophysical law maps this physical-functional state onto the distinct state of standing in an irreducible relation of conscious representation to edenic red, which is precisely what is needed for you to have a basic *reason* to believe *I'm conscious of the quality red*. Likewise, the physical-functional states of your

colour system cause you to make certain similarity-reports and engage in certain ordering behaviour. The psychological laws map them onto distinct colour experiences whose reason-grounding roles are harmonious with these responses.

On epiphenomenalist dualism, normative harmony might have failed. For instance, different laws might instead have mapped all our physical-functional states P_1, P_2, … onto the *same* bodily sensation: a full-body lukewarm sensation (as when you are in lukewarm pool). In this scenario, we're blind as bats and always have the same simple sensation — we're nearly zombies. Yet, given causal closure, all of our complex behaviours and responses would be the same, and therefore totally out of whack with what we have reason to believe and desire given our simple experience. What — short of an intelligent designer — might explain why the psychological laws are actually 'fine-tuned' to result in normative harmony?

Here is another illustration. Suppose that neural pattern Z is caused by eating rotten flesh and causes withdrawal. Neural pattern X is caused by eating oranges and causes eating more. These are *affective responses*. Now, it is in the essence of the experience of edenic rottenness to provide a reason to *desire* that the experience stop, and it is in the essence of the experience of edenic sweetness to provide a reason to *desire* that the experience continue. That is, the first experience is essentially bad (disvalue) while the second is essentially good (value). In accordance with normative harmony, the psychophysical laws map pattern Z and pattern X onto these experiences, resulting in a harmony between reasons and behaviour. But, in a physically identical world, the psychological laws could have mapped Z and X onto *valence-inverted* experiences — the kind of possibility Chalmers alludes to in the passage quoted above. In this horrific *yuck-yum inversion scenario*, oranges taste like rotten flesh but (given causal closure) we gobble them up and say 'they taste great' just as in the actual world.[2] Why does normative harmony actually obtain?

Notice that causal closure is a crucial element of the normative harmony problem. Interactionist dualists who reject causal closure might explain normative harmony by saying that our behavioural

---

[2]   These subjects exhibit a kind of rational absurdity. But we should not confuse rational absurdity with metaphysical impossibility. The cases are certainly conceivable (i.e. cannot be conclusively ruled out *a priori*).

responses are sensitive, in a top-down way, to the built-in normative features of our experiences (see Saad, 2019).

Now that we understand the normative harmony problem, we can see that it arises equally for an 'emergentist' form of physicalism that endorses distinctness, normative significance, and causal closure:

Conscious state C ⟶ Reasons for belief and desire

Necessary ↑   ⤡ ||?

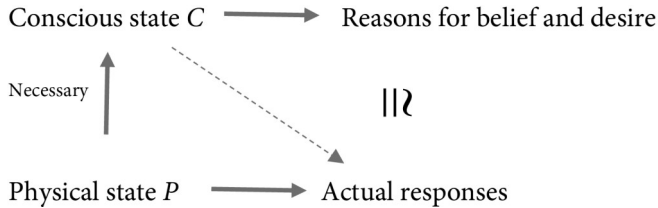Physical state P ⟶ Actual responses

Figure 2. Emergent physicalism

Emergent physicalism is just like epiphenomenalist dualism. In fact, it might accept the very same psychophysical laws of form [#]. The only difference: it maintains that those same laws have a different modal status, in particular, they are metaphysically necessary 'grounding laws', rather than contingent laws of nature (Schaffer, forthcoming).

Emergent physicalism is stronger than 'non-reductive physicalism'. Non-reductive physicalists deny that experience properties are identical with 'physical properties' but they often hold that they are identical with physically-realized 'functional properties'. By contrast, emergent physicalism holds that experience properties are even distinct from 'functional properties'. In fact, in one version, they are *quite different* from the underlying physical-functional properties in their structure and normative nature: they consist in standing in an irreducible relation of conscious representation to edenic qualities, and it is in their constitutive essence to provide reasons for certain beliefs and desires.

As I said, on emergent physicalism, conscious experiences of edenic qualities (C) are *grounded by* the computational physical states (P) that causally explain our consciousness-related responses, rather than being merely *nomically determined by* them as on epiphenomenalist dualism. Given a suitable counterfactual/proportionality theory of causation, this means that they count as 'supervenient causes' of our responses. In Figure 2, this is indicated by the dotted arrow.

Here is why emergent physicalism faces the normative harmony problem no less than epiphenomenalist dualism. True, it is unlike epiphenomenalist dualism in holding that laws of form [#] are

metaphysically necessary grounding laws. But it still implies norma-tive harmony. That is, it still implies a striking pattern in the 'psycho-physical grounding laws': they always correlate a physical state *P* with a distinct conscious experience *C* whose essential normative role in providing reasons is *harmonious* with the causal role of *P* in generating responses, in the ways I indicated above. Further, since the basic psychophysical grounding laws have no explanation and are independent of the normative laws, the agreement here would seem inexplicable. It must be accepted with 'natural piety'. This should lower our credence in the view.

There were already familiar problems with emergent physicalism. Proponents of this view accept laws of form [#] — the very same laws connecting distinct physical and phenomenal states accepted by dual-ists. True, they hold that these laws are metaphysically necessary grounding laws, rather than contingent laws as dualists maintain. But this does not change the fact that they add to the complexity of our theory of the world (unlike psychophysical identities). And they are anomalous in that they are quite different from all other general fundamental-to-non-fundamental grounding principles (e.g. mereol-ogical principles, principles of complex-property-formation). The normative harmony problem adds to these familiar problems. It shows that the grounding laws have a further implausible feature: in each and every case they inexplicably always satisfy normative harmony.[3]

In §2, I rejected Chalmers' suggestion that modal independence alone is *sufficient for* a problem about luck or coincidence. Now we see that modal independence is not *necessary*. For emergent physicalists can reject modal independence. But they still face a coincidence problem because they still accept normative harmony, as I have explained.

---

[3]   To see clearly that the normative problem goes beyond the familiar problems, consider Campbell's (1993) 'emergent physicalism' about *colours*. It faces the familiar problems — it requires anomalous, complex grounding laws between reflectances and edenic colours — but it *doesn't* face the normative harmony problem. It is not the case that, of all the epistemically possible chromatic-physical associations ('chromatic grounding laws'), the one that in fact obtains stands out as achieving 'normative harmony'. For, unlike experiences, edenic colours are normatively inert.

4. Response: Identity Theory?

What is the solution to the normative harmony problem?

Notice that the very formulation of normative harmony builds in a commitment to both *distinctness* and *essential normative significance.* It requires the kind of picture of conscious experience represented in Figures 1 and 2. There are two things: the causal roles of our physical states (the bottom level), and the essential normative roles of our experiences (the top level). The problem concerns the coincidence between the two.

Therefore, one way to avoid the problem would be to adopt a different picture of consciousness that rejects these commitments. For example, let us again look at the *identity theory* discussed in §2. Identity theorists reject *distinctness*. It is not the case that your conscious experience is a relation to edenic qualities that is distinct from, but dependent on, your neural state P, which cases your responses. Rather, it is *nothing but* your neural state P.

Physical state P ⟶ Actual responses

Figure 3. Identity theory

Identity theorists also reject *essential normative significance*. On their view, colour experiences, taste experiences, experiences of pain and pleasure are nothing but neural patterns. And it is very hard to see how it might be in the essences of particular neural patterns that they provide *reasons* for certain *beliefs* or *desires*. Instead, identity theorists are likely to accept a *reliabilist* reduction of the justificatory role of experiences (neural states) and a *response-dependent*, *desire-based* reduction of their goodness and badness. It's not in the *essences* of any neural patterns to reliably cause certain beliefs or to be desired by normal subjects. So, on these views, it's not in their *essences* to justify certain beliefs, or to be good or bad. True, pre-theoretically, it is very natural to think that some experiences *essentially* provide reasons for certain beliefs and some experiences are *essentially* good or bad. But the present view denies this. The normative features of experiences derive from their non-essential causal roles.[4]

---

4   Identity theorists who take a response-dependent view of valence might appeal to 'rigidification' on our actual responses in order to provide true readings to statements like 'Necessarily, pain (on this view, a neural pattern) is bad'. But they still cannot claim that it flows from the *essence of pain itself* (on this view, a certain neural pattern)

Since the formulation of normative harmony builds in a commit-
ment to distinctness and essential normative significance, and since
identity theorists reject these assumptions, they reject normative
harmony *as I have formulated it*. And since they reject it, they face no
problem about its explanation. In other words, the normative harmony
problem is about the coincidence between 'two things' — the causal
roles of our physical states and the essential normative roles of our
distinct experiences. Identity theorists deny that there are such 'two
things' here, thereby dodging this particular problem.

However, maybe identity theorists face a problem in the vicinity.
For instance, they do accept the following regularities: (I) experiences
tend to cause doxastic responses (introspective and perceptual) that
are 'fitting' or 'justified' and (II) we tend to act as if we desire good
sensations and act if we don't desire bad sensations. What explains (I)
and (II)?

Identity theorists might respond that (I) holds because experiences
(neural states) are likely to be connected to reliable (introspective and
perceptual) belief-forming mechanisms and the outputs of those are
ipso facto justified. And they might say that (II) holds because of the
response-dependent nature of the goodness and badness of sensations.
Crudely, a *good sensation* (that is, on the identity theory, a 'good'
neural state) is just one we desire to have, and in turn *desiring to have*
is (roughly) acting as if one desires to have; whereas a *bad sensation*
(that is, a 'bad' neural state) is just one we desire not to have and
desiring not to have is acting as if one desires not to have (being
disposed to withdraw, etc.). Putting these together, identity theorists
can explain (II): why we tend to act as if we desire good sensations
and act as if we don't desire bad sensations.

Another attempt to saddle identity theorists with a troubling
regularity looks for a pattern in the psychophysical identities. For
example: [R] *taste sensations (e.g. sweet-citrus experience, rotten-
flesh experience) with their essential reason-grounding roles always
happen to be identical with physical states (X-pattern, Y-pattern) that
cause responses matching those essential reason-grounding roles*.
Now, these identities are conceptually independent from the physical
facts; it is *conceptually* possible that they should instead have been
such as to falsify [R] — even *inverted*. Is regularity [R], and this sense

---

that it is bad. (Perhaps the essentialist claim could be accommodated by a *functionalist*
identity theory of experiences that builds in our typical responses.)

of 'could have been otherwise', enough to generate a luck problem for identity theorists? Could a problem for them be that the *pattern of identities* is lucky?

Identity theorists can reply that the alleged regularity [R] pre-supposes that experiences have essential reason-grounding roles. But, as we saw, they can reject this. In that case, [R] cannot be troubling because it is false.

In short, the identity theory of conscious experiences, when supple-mented with a reliability theory of their justifying role and response-dependent theory of their goodness and badness, might avoid any problem about 'luck' or 'coincidence'.

Still, I cannot accept such an identity theory — and I don't think Chalmers can either. Experiences are relations to edenic colours and shapes and so on. But this is not true of the underlying neural patterns, computational states, and so on: they can be characterized fully with-out mentioning edenic colours and shapes. In addition, it's con-ceivable that a mereologically simple thing should have an experi-ence; but such a thing couldn't undergo a neural pattern. For these reasons, experiences must be distinct from underlying neurocomputa-tional states, even they are dependent on them.[5] In addition, I think it's just obvious that some experiences *essentially* provide reasons to believe certain things (contrary to reliabilism), and that some experi-ences are *essentially* good or bad (contrary to response-dependent theories of valence).

So, *if we have conscious experiences at all*, the right view will be either dualism (Figure 1) or emergent physicalism (Figure 2) — resulting in the normative harmony problem.

5. Response: Strong Illusionism?

This suggests a truly desperate solution: *maybe we don't have con-scious experiences at all!* For example, we are familiar with the idea that items possessing colours-as-we-see-them (edenic colours) are illusions. Why not go one step further and say that it's also an illusion that we have *experiences* as of colours?

My first reaction is the same as Chalmers' (2018, p. 54). In some cases, it is *certain that I have an experience*.

---

5   These arguments work equally against a 'Russellian monist' form of the identity theory (see Pautz, 2015, Section 2).

But I worry about this. For what could *explain* our having indefeasible justification for certain introspective beliefs (I'm having an experience), but defeasible justification for perceptual beliefs (there is a reddish item there)?

One explanation — favoured by Chalmers (*ibid.*, p. 25) — is that (i) here in the actual world we are automatically 'acquainted' with our own states of consciousness but not the states of the mind-independent world and (ii) acquaintance suffices for indefeasible justification. But (ii) faces a simple objection. Presumably, in some *other* 'possible world', we might be acquainted with an *external* state, say the redness of a tomato; but here acquaintance would not yield indefeasible justification (if in this world you were presented with strong but misleading arguments for colour illusionism you should decrease your credence in the claim that a red thing is there).

Another explanation is that all and only phenomenal truths about yourself are automatically part of your evidence (the 'phenomenal conception of evidence'), so that you trivially have maximal justification for believing all such truths (they have probability 1 given your evidence). But this explanation overshoots: it entails that you even have maximal justification for believing obscure structural truths about experiences (e.g. the ratio scale for loudness). This is contrary to my (tenuous) grasp of the term-of-art 'propositional justification'.

Maybe the lesson is that we can be certain that we are conscious but this is an *inexplicable* epistemic fact. But suppose instead that we must give up on certainty. We merely have *defeasible* reason to think we are conscious. Even then I think we should reject illusionism.

The fact is, illusionism isn't helpful. Arguably, no reductive theory of representation could explain how we have illusions in which we (falsely) represent ourselves as *being conscious of edenic red*, if nothing has ever been conscious of edenic red and these terms ('conscious of', 'edenic red') are indefinable. If so, illusionists require a problematic form of emergent physicalism (§3) on which the representation relation here is irreducible to any physical (e.g. causal-informational) relation.[6] Additionally, if the illusion states essentially play certain normative roles (e.g. it's in the nature of the illusion of

---

[6] The representational problem is: how can one represent that one *is conscious*, if no one has been conscious and what it is to be conscious is indefinable? This formulation of the problem is neutral concerning whether 'is conscious' expresses an uninstantiated property (Chalmers, 2018, p. 26); it arises equally for illusionists who accept nominalism.

pain that we have a reason to avoid it; see Kammerer, forthcoming, for discussion) and those normative roles are harmonious with our responses, it may face a reformulated version of the normative harmony problem, applied to illusions of experiences rather than to experiences.

6. Conclusion

I don't have a good response to the normative harmony problem (but see Pautz, 2010, p. 358, for a desperate attempt). I consider it to be a deeply troubling but largely overlooked element of the mind–body problem.

References

Campbell, J. (1993) A simple view of colour, in Haldane, J. & Wright, C. (eds.) *Reality: Representation and Projection*, Oxford: Oxford University Press.

Chalmers, D.J. (2018) The meta-problem of consciousness, *Journal of Consciousness Studies*, **25** (9–10), pp. 6–61.

Kammerer, F. (forthcoming) The normative challenge for illusionist views of consciousness, *Ergo*.

Lee, G. (2019) Alien subjectivity and the importance of consciousness, in Pautz, A. & Stoljar, D. (eds.) *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness*, Cambridge, MA: MIT Press.

Pautz, A. (2010) Do theories of consciousness rest on a mistake?, *Philosophical Issues*, **20**, pp. 333–367.

Pautz, A. (2015) A dilemma for Russellian monists about consciousness, [Online], https://philpapers.org/archive/PAUCRM.pdf [18 August 2019].

Saad, B. (2019) A teleological strategy for solving the meta-problem of consciousness, *Journal of Consciousness Studies*, **26** (9–10), pp. 205–216.

Schaffer, J. (forthcoming) Ground functionalism, *Oxford Studies in the Philosophy of Mind*.

White, R. (2005) Explanation as a guide to induction, *Philosopher's Imprint*, **5**, pp. 1–29.