



Self and Knowledge in Psychedelic Therapy Reply to Commentaries on *Philosophy of Psychedelics*

Chris Letheby^{a,b}  (chris.letheby@uwa.edu.au)

Abstract

Much of my book *Philosophy of Psychedelics* is devoted to elaborating and defending two basic claims: that psychedelic therapy works mainly by changing mental representations of the self, and that it has many epistemic benefits consistent with a naturalistic worldview. The commentaries in this symposium generally focus on one or the other of these claims. On the mechanistic front, the commentaries by Hoffman and by Martin and Sterzer seek to supplement my account by drawing attention to factors it does not emphasize. In this reply I agree with most of their claims, but propose some important qualifications. Meanwhile, the commentaries by Lyon and Farennikova and by Colombo challenge some core commitments of my account. I agree with many of their claims, but defend some of these core commitments. On the epistemological front, Bortolotti and Murphy-Hollies endorse the overall shape of my account and make some productive connections to ideas concerning agency-first epistemology and self know-how. Caporuscio brings into play the idea of self-shaping, which connects to self know-how, but argues against certain core claims of my account. I endorse many of these claims about agency-first, self know-how, and self-shaping, but propose some important qualifications. Finally, Fink agrees that psychedelic experiences can have some epistemic benefits, but argues that these do not pertain to epistemic justification, and that the main benefit is increased understanding. I am sympathetic to Fink's proposal about understanding, but criticize some of his arguments for the thesis of Psychedelic Justification Impossibilism.

Keywords

Ayahuasca · DMT · Epistemic agency · Epistemic justification · Epistemology · Hallucinogen · LSD · Mescaline · Philosophy · Predictive processing · Psilocybin · Psychedelics · Self know-how · Self-representation · Self-shaping · Understanding

This article is part of a symposium on Chris Letheby's book "Philosophy of Psychedelics" (OUP 2021), edited by Chiara Caporuscio and Sascha Benjamin Fink.

^aThe University of Western Australia.

^bThe University of Adelaide.

1 Introduction

I would like to begin by thanking all the contributors to this symposium for their generous and constructive engagement with the ideas presented in *Philosophy of Psychedelics* (henceforth POP, Letheby, 2021). Time constraints inherent to the production of a symposium such as this have inevitably limited my reflection on their arguments. To paraphrase Bilbo Baggins from Tolkien's *Lord of the Rings*, I have thought about less than half of these arguments half as deeply as I should like, and here I respond to less than half of them half as adequately as they deserve.

The bulk of POP is devoted to elaborating and defending two basic claims: that psychedelic therapy works mainly by changing mental representations of the self, and that it has many epistemic benefits consistent with a naturalistic worldview. In general, the commentaries address one or the other of these claims¹; accordingly, I have structured this reply into two main sections. In the interests of brevity, I do not recapitulate the commentators' arguments in great detail, except where there are substantial points of disagreement. I assume that readers of this reply have read the book (or the précis) and the commentaries.

2 Mechanisms

The commentaries I discuss in this section vary in how strongly they challenge my claims about the causal mechanisms of psychedelic therapy. The commentaries by Hoffman and by Martin and Sterzer seek to supplement my account by drawing attention to factors it does not emphasize. Meanwhile, the commentaries by Lyon and Farennikova and by Colombo challenge core commitments of my account, to one degree or another.

2.1 Positive affect

Hoffman (2022) endorses the general shape of the account of psychedelic therapy sketched in POP, but argues that it does not sufficiently emphasize the role of positive affect, and that a complete account will need to do so. Here, I am in agreement. She also argues that exploring the role of affect calls into question some of my claims concerning ego dissolution; on this point, I am less convinced.

It is true that I do not strongly emphasize or deeply explore the affective dimensions of psychedelic therapy in POP, an omission that I briefly acknowledge (Letheby, 2021, p. 202). Hoffman cites several studies linking affective factors to clinical outcomes. These findings suggest that a complete theory of psychedelic therapy must account for affective factors. As Hoffman notes, the plausible link between affect and self-representation means this observation is consistent with

¹The exception is Johnson's (2022) wonderful introduction, 'Psychedelic science needs philosophy', with which – of course! – I have no disagreements of note.

the overall shape of my account: the account needs to be expanded to include affective factors, not fundamentally revised. Thus far, I agree entirely.

Hoffman explores the role of affective factors via a comparison of psychedelic therapy and MDMA-assisted therapy, arguing that this comparison reveals some potential weaknesses in my account. I hold that psychedelics, by disrupting self-binding processes, move the phenomenal self-model along the continuum from phenomenal transparency to opacity. Hoffman comments that “this way of describing things requires a commitment to the predictive processing [PP] account, which opens the theory to objections” (Hoffman, 2022). But I do not think that it does require such a commitment. The idea that transparent phenomenal self-modelling results from a binding process, and that psychedelic-induced unbinding renders self-modelling opaque, is independent of PP. PP is one way of spelling out the details – but the self-binding theory of self-representation was developed independently of PP (Sui & Humphreys, 2015) and internalist “controlled hallucination” views of phenomenal consciousness have been, too (e.g. Revonsuo, 2006). PP is one influential basis for this kind of view of phenomenal consciousness and self-consciousness, but the general view can be argued for independently of PP (Metzinger, 2004; Revonsuo, 2006) – indeed, I think psychedelic experience tends to support it.

Hoffman also says that my account of the role of phenomenal opacity in psychedelic therapy “hides an ambiguity between cognitive and phenomenal opacity” (Hoffman, 2022). While the term “cognitive opacity” is Hoffman’s, the distinction originates in Metzinger (2004). The distinction is as follows: while phenomenal opacity involves *experiencing* something as a representation, cognitive opacity involves merely *thinking*, or *judging*, that something is a representation (Hoffman, personal communication). Invoking this distinction, Hoffman comments that “metacognitions about the self will always make those parts of the self-model that are their objects more opaque *but not typically phenomenally opaque*” (Hoffman, 2022). Thus, she contends, while self-model opacity of some kind is plausibly involved in psychedelic therapy, it may not always be *phenomenal* opacity.

Perhaps this is correct; perhaps psychedelic therapy sometimes involve merely cognitive opacity of the self-model. It still seems to me that there is an important difference between experientially-backed cognitive opacity and the kind one can get from, e.g., book-learning. A and B may both *believe* that their sense of self is continually constructed by a mental modelling process, and is therefore malleable. But if A’s belief results from reading philosophy and watching Anil Seth’s TED talk, while B’s results from a psychedelic experience of self-unbinding, then B’s is likely, *ceteris paribus*, to carry deeper conviction and have a greater existential and behavioural impact. Of course, this conjecture could be tested empirically, if we could find a way to measure these different kinds of opacity.

I also think that mere cognitive opacity, and full-blown phenomenal opacity, can *both* result from disruption to predictive self-binding. This hypothesis is not

undermined by the observation that MDMA (a) seems to induce therapeutic effects via some kind of self-unbinding, but (b) rarely, if ever, induces full-blown mystical experiences, ego dissolution, and global phenomenal opacity. This observation is consistent with the idea that disruption to self-binding is key to the therapeutic effects of both MDMA and psychedelics, even if the latter can induce types and degrees of unbinding that the former rarely does (i.e. disrupt more aspects of the self-model, and to greater extents). Despite the similarities Hoffman notes, there are important differences between not just psychedelics and MDMA, but psychedelic-assisted and MDMA-assisted *therapy*. For instance, Hoffman's description of "inducing a blissful state to facilitate productive engagement with painful material" (Hoffman, 2022) may be a fair description of MDMA's therapeutic mechanisms, but it is less clear that it adequately characterizes typical therapeutic *psychedelic* experiences. Notwithstanding the undeniable role of positive valence in the latter, many parts of these experiences are far from blissful. In some cases an undesired but unavoidable confrontation with painful material is involved, and seems to be facilitated not by a blissful state but by a diminished capacity to avoid engaging with uncomfortable mental contents – plausibly a disruption to self-binding, but not an increase in positive affect.

In sum, positive affect is an important factor in psychedelic therapy that is underexplored in POP, and comparing and contrasting psychedelic-assisted and MDMA-assisted therapy seems like a fruitful direction for future research. But total ego dissolution may still be the limit case of an unbinding mechanism that underlies the therapeutic effects of both substances – even if taking this mechanism to its limits is not therapeutically necessary. In section 2.4 below I formulate a distinctive empirical prediction of the self-binding account of psychedelic therapy; perhaps formulating predictions of this kind could allow us to contrastively test some of Hoffman's and my claims about similarities and differences between psychedelic- and MDMA-assisted therapy².

2.2 Context and strengthened beliefs

Like Hoffman, Martin and Sterzer (M&S) (2022) seek to supplement my account with attention to under-emphasized factors, not to challenge its core tenets. Here the under-emphasized factors are (a) the role of strengthened priors and (b) the influence of context on the kinds of alternative self-models that patients encounter during therapeutic experiences. I am almost entirely in agreement with M&S, and have little to add to their discussion. As they note, the importance of strengthened priors can be made compatible with the REBUS (RElaxed Beliefs Under pSychedelics) model, "so long as these strong prior influences emerge at lower levels in response to the relaxed constraining influence of priors at the highest levels of the system [...] the REBUS model may involve both 'strong'

²I am indebted to Matteo Colombo for raising this point.

and ‘weak’ prior influences, which are present at different hierarchical levels to different degrees” (Martin & Sterzer, 2022).

I think this is the correct line to take. It is true that psychedelics strengthen some priors, as well as weakening others (Safron, 2020); that this is important to understanding their experiential and therapeutic effects; and that they can be “generally characterised as weakening the constraining influence of priors encoding rigid highly-abstract beliefs of the world and self, and (relatively) strengthening prior beliefs associated with contextual sources, leading to an amplified influence of ‘set’ and ‘setting’ on the mental state of the user” (Martin & Sterzer, 2022). I also agree that the account I develop under-emphasizes strengthened priors, and that investigating this factor in more detail can help fill out the account by accounting for the crucial role of context in psychedelic therapy. The cognitive playing field is, indeed, not *totally* level in the psychedelic experience or in the post-acute phase of psychedelic therapy, and something must explain why certain specific (often healthier) forms of self-modelling tend to emerge, rather than others. Surely, as M&S say, therapeutic expectations, intentions, and settings will figure importantly in the explanation.

I would add, however, that these may not be the only important factors. There are many anecdotal accounts of people discovering and consolidating healthier forms of self-modelling after psychedelic experiences that lacked any therapeutic intent or setting. In this light, consider that many patients do not describe discovering new forms of self-modelling, but rediscovering old, forgotten, or disused forms. As one patient put it: “At [its] most basic, I feel like I used to before the depression” (Watts, Day, Krzanowski, Nutt, & Carhart-Harris, 2017, p. 531). This idea is expressed by in neurobiological terms by Nichols et al.: “In essence, the brain is able to restore its network connectivity to a predisease state” (2017, p. 215). This suggests another factor that may explain why specific forms of self-modelling emerge, at least in some cases: they are old, familiar forms that are the most probable alternatives available to the brain when the recent, default model is relaxed. Of course, set and setting undeniably play a large role in such cases (as Nichols et al. emphasize). But another important factor – which can be subsumed under “set” in a broad sense – consists simply of the ready availability of certain cognitive “materials” (i.e. models) to the brain, based on extensive prior experience.

When considering the relationship between therapeutic intentions and therapeutic outcomes, another factor should be kept in mind. This is that some patients’ reports draw attention not to the development of specific new self-models (“being a non-smoker”) but to a loosening of the grip of self-modelling on attention and salience altogether – creating a more flexible and less tyrannical self-model, rather than one with specific contents. In the qualitative study of psilocybin-assisted smoking cessation that M&S discuss, quitting smoking “was often reported as one of the *least* important effects of the study for participants” (Noorani, Garcia-Romeu, Swift, Griffiths, & Johnson, 2018, p. 763). This is exemplified by multiple reports; for example:

This is about a smoking study, I keep forgetting that. Because there's so much more that happened [...] [Smoking] just seems so petty compared to some of the stuff that was happening. [Participant 405]

It was all about searching for answers to questions that are age-old. Maybe we have the answer to some of it, maybe we'll never have the answer to it. But none of it had to do with addiction to cigarettes. It all had to do with stretching space and time, and asking questions like, 'Why is there something rather than nothing?' And, 'What happened before the Big Bang?' [...] All those things that had nothing, absolutely nothing – at least in the conscious thinking of it – with stopping smoking [...] every time I think of a cigarette, it brings me back to the three sessions of the psilocybin trip, where you're trying to ask questions that there may not be answers to. So, that's more fun than smoking! [Participant 422] (Noorani et al., 2018, p. 763)

Reports like this suggest a strong role for disidentification with existing self-narratives, and disruption to the patterns of attention and valuation that they underpin, as opposed to the strengthening of specific new (or old) self-narratives. Clearly a complex interplay of factors is at work. In sum, however, I agree with M&S: strengthened priors are important, and ought to be taken into account; they can help to explain important contextual influences in psychedelic therapy; and their main epistemic benefits probably amount to new knowledge of old facts.

2.3 Transparency and opacity

While endorsing many claims of my account, Lyon and Farennikova (L&F) (2022) argue against one of its central tenets. They argue that: (a) my account says increased phenomenal opacity is more important in psychedelic therapy than increased transparency; but (b) increased transparency is at least as important, if not more so; therefore (c) a core claim of my account is false. They take the many different ways of drawing the opacity/transparency distinction to be essentially equivalent. They describe five types of cases in which, they say, psychedelics increase transparency with plausible therapeutic import: (i) increased transparency in representations of nature, (ii) increased transparency in representations of music, (iii) increased transparency of mental imagery, (iv) an amplified sense of reality (e.g. the "noetic quality"), and (v) apparent revelations of a truer or more authentic self.

I am not sure these are all cases of increased transparency, as I define that term. I deliberately stick to a definition of transparency as *not experiencing representations as representations*, since it is unclear whether this phenomenal property results from certain postulated mechanisms, such as the attentional unavailability of earlier processing stages (Metzinger, 2004). And while there are important links between transparency/opacity and other dichotomies, such as a sense of reality/unreality, unity/duality, etc., I do not think these links amount to equivalences.

The link between transparency/opacity and reality/unreality does, in my view. But I am less convinced that transparency/opacity maps neatly, rather than approximately, onto unity/duality, absorption/disconnection, necessity/optionality, etc.

L&F's analysis presupposes such a neat mapping, and bidirectional entailments between these various dichotomies. Their first type of case involves an experiential sense of connection or unity with nature. I think these sorts of cases plausibly involve increased transparency, but not necessarily of one's representations of nature. Rather, what becomes more transparent is one's representations of *one's unity with, or connectedness to, nature*. People do not typically describe nature itself seeming more real, but their connection to it seeming more real. In fact, only one of L&F's examples talks about anything seeming more real, as opposed to simply about experiencing unity/connectedness with nature: "All kinds of other things were coming, too, like feelings of being connected to everything [...] and it wasn't like talking about it, which makes it an idea, it was, like, experiential." Cases like this fit with my overall account: psychedelics (a) disintegrate our default self-models, (b) allowing healthier forms of self-modelling to be discovered and subsequently consolidated.

L&F's second type of case involves music. They say it is clear that these kinds of cases involve increases in transparency – but on my definition of transparency, this is not made clear by the quoted reports. None of these talk explicitly about music seeming more real, or less like a representation. Rather, they talk about increased feelings of *unity, connection, and absorption* with/in the music, and, in some cases, the music seeming to take on different properties, such as agency or being the vehicle for various experiences/realizations. These experiences *may* feature increased transparency (in my sense) of subjects' mental representations of music, but the quoted reports do not say as much.

Similar points apply to L&F's third type of case: mental imagery. They quote the findings of a qualitative study that the "visuals" experienced by psychedelic subjects often serve as "principle organizing motifs of subjective experience with multifold vectors (e.g., audiovisual, relational, autobiographical, spiritual, epistemological, ontological) [and are] often deeply coded with layers of meaning by participants." They note that subjects often become deeply absorbed in this imagery. But none of this indicates that subjects do not experience their mental imagery as representational. In some cases, psychedelic visuals certainly do have a high degree of transparency. For example, high-dose DMT subjects often report visiting immersive other worlds that feel "more real than real" (Gallimore & Strassman, 2016). But it seems equally likely that some participants undergo rich and vivid visual experiences that are deeply meaningful and absorbing, while simultaneously recognizing these experiences to be drug-influenced mental events. The fact that imagery is experienced as vivid, meaningful, important, and absorbing does not entail that it is not experienced as representational (compare artworks of various kinds).

L&F's fourth and fifth types of cases seem to me the clearest and most plausible. The fourth involves the noetic quality which is criterial of the mystical-type

experience – the feeling of gaining knowledge of ultimate reality. This undeniably involves increased phenomenal transparency as I define it. Likewise with the fifth type of case, which involves seeming to discover a better, deeper, or more authentic self. This sort of experience involves increased phenomenal transparency, and is often central to the therapeutic process.

However, before one can discover a better, deeper, or more authentic self, one must let go of a worse, shallower, or less authentic self. Thus, L&F's fifth type of case fits well with the mechanism that I posit: unbinding and increased opacity of the default self-model, leading to the discovery and subsequent consolidation of different, healthier forms of self-modelling.

L&F anticipate this line of reasoning. They respond that even if transparency increases causally depend on prior opacity increases, this does not mean that the latter are more important. They use the analogy of someone who travels to another country for a relaxing holiday. Typically, such a journey causally depends on the prior purchase of a plane ticket. But this does not mean that the purchase of the ticket explains why the person is more relaxed: to explain this, we must look to the experiences they had during the trip.

I think this analogy is suggestive, but can be adapted illuminatingly. Rather than relaxing holidays, consider a case even closer to our present concerns: *transformational* travel. Imagine several people, each of whom undertakes a meaningful, transformative journey to another country. Each person goes to a different country, has different experiences, and is transformed in different (but all salutary) ways. Is there a common mechanism? Is there any robust, informative generalization we can make? Clearly we cannot say things like “they went to Australia and hitch-hiked across the Nullarbor Plain” if only one of them did this, while another went to Nepal and undertook a meditation retreat, and yet another went to France, worked in a vineyard, and had a whirlwind romance. However, it clearly would be equally unhelpful to say that each person bought a plane ticket. I think the most informative generalization would be something like this: each person left their home, and with it their normal modes of thought, feeling, and action; each went on a transformative journey to other places in which they had rich, meaningful, and unfamiliar experiences. Despite the varying details, each returned home with an expanded sense of possibilities and a vivid, experientially-backed sense of the contingency of their ordinary modes of living and experiencing. Each left their ordinary home and ordinary routines, went on a transformative trip, and returned with the profound conviction that ‘there are other ways I can be’ – the conviction, most basically and generally, that ‘things [and I] *can be otherwise*’.

Of course, the negative insight that one's ordinary ways of being are contingent, and the positive but abstract insight that one can be otherwise, are empty without some definite acquaintance with specific, alternative ways of being. But if we want to generalize robustly and informatively, then we need to strike the right level of abstraction. And the details of the specific forms of self-modelling that psychedelics disrupt, and of the specific alternative forms of self-modelling

that they can make available, are too heterogeneous to permit much informative generalization at this level. What I think we can say of nearly all therapeutic psychedelic experiences is that people experience different ways of seeing and being, and (equivalently, in my view) different forms of self-modelling, due to the unbinding of their default self-model; that they gain a deep, experientially-backed conviction of the possibility of change and the contingency of their default sense of self; and that these components interact to cause lasting change. True, the discovery of new forms of self-modelling is just as important as the discovery of the contingency of old forms, and I emphasize the latter; but the latter is, nonetheless, typical, generalizable, and important.

However, I think L&F are right about one important point: The phenomenon that I call “increased opacity of the self-model” is *not* always that, given my understanding of opacity. In essence, I make the same sort of mistake with which I earlier charged L&F: I assume that because self-model opacity tends to cause insight into the mutability of one’s self, any insight into the mutability of one’s self must amount to, or result from, self-model opacity. Here I was blinded by my relatively idiosyncratic theoretical conviction that there are no selves apart from self-models. In my worldview, a distinction between *the mutability of the self* and *the mutability of the self-model* is a distinction without a difference. But to many people who do not share my worldview, the phenomenon I am talking about may be at least as likely to present in the former guise as in the latter.

Put simply, the objective *fact* that subjects discover – according to me – is that their ordinary sense of self is malleable, in virtue of its being continuously constructed by a mental modelling process. But often, the way subjects conceive of that fact probably has little to do with self-modelling or self-representation, and more to do with the possibility of change for the “self” or the person. The slogan version of the insight, as it presents to many subjects, is “I can be otherwise” – not “my self-model can be otherwise”.

2.4 Serotonin and predictive processing

Colombo (2022) argues that inconsistency in the PP literature about the neural implementation of precision weighting and the computational functions of serotonin calls into question the reliance of my account on the PP framework. He notes that other computational frameworks for the explanation of psychiatric phenomena exist, and argues that one such framework, reinforcement learning (RL), promises to provide plausible alternative explanations of psychedelic therapy that do a better job of linking different levels of description.

Actually, I think my main claims are independent of PP. Here are some mechanistic claims I make in the book, in descending order of my confidence in them based on current evidence: (1) psychedelic therapy works mainly by changing mental representations of the self; (2) a central function of these representations is to bind self-relevant information across domains and levels of processing; and (3)

this binding function is performed according to PP principles. In my view, the current evidence concerning psychedelic therapy and self-representation makes quite a strong case for (1). The evidence surveyed by Sui and Humphreys (2015) seems to provide a fairly strong case for (2). And (3) is one speculative sketch of how this might work. The truth of PP has not been established definitively, and it may turn out that self-representation (and self-binding) do not operate according to PP principles. Thus, the specific details of my account should be held fairly lightly, and subjected to scrutiny, such as Colombo's. However, this does not impugn the overarching claims of the account, unless it turns out that there is *no* plausible way to spell out its details. In essence, I think Colombo's arguments importantly challenge the details, but not the core, of my account.

Colombo anticipates this line of response. Having sketched two possible RL explanations of psychedelic therapy, he comments:

One possible [response] is that the truth of either of these RL hypotheses would *not* change anything in Letheby's key idea [...] Yet, if different computational hypotheses about serotonin are all compatible with Letheby's idea that psychedelic therapy mainly works by changing self-representations, then one may conclude that this idea is too generic to advance existing understanding of psychedelic therapy, or of serotonin function in relation to self-representations. (Colombo, 2022).

I concede that the bare idea that psychedelic therapy mainly works by changing self-representations is too generic to advance existing understanding of serotonin function in relation to self-representations. For it to do this, Colombo is right: one needs more specific neurocomputational hypotheses, fleshed out with more implementation detail.

However, my main goal was not to advance understanding of serotonin function in relation to self-representations – it was to advance understanding of psychedelic therapy. And I do not think the self-representation hypothesis is too generic to do that, even if it is compatible with different computational hypotheses about serotonin. The self-representation hypothesis is a specific, testable proposal about how psychedelic therapy works. It makes different predictions to the Molecular Neuroplasticity Theory, to the Metaphysical Belief and Alief theories, and to other prominent explanatory hypotheses (e.g. Hendricks, 2018). One might worry that the self-representation hypothesis and the Molecular Neuroplasticity Theory are pitched at different levels of explanation³, and therefore cannot really be in competition. However, in chapter 4 of POP I argue that at least one version of the Molecular Neuroplasticity theory – the *Pure Neuroplasticity Theory* – is genuinely inconsistent with (neuro) cognitive and psychological theories that appeal to changes in self-representation, metaphysical beliefs, and so forth.

³Perhaps cognitive vs. neural, psychological vs. neuroscientific, or algorithmic vs. implementational, depending on one's framework and assumptions.

When we flesh the self-representation hypothesis out with the self-binding theory, while it remains compatible with multiple implementational hypotheses, it becomes even more specific and testable. Let me spell out one prediction of a self-binding account of psychedelic therapy. Sui and Humphreys (2015) reported on an experiment (Wang, Humphreys, & Sui, 2016) in which participants were asked to associate neutral shapes with themselves and with other people (friends and strangers). Later, they were asked to re-associate the same shapes with different people. For shapes that had been linked to the self, this proved difficult – there was a switching cost. Sui and Humphreys (2015) interpret this in terms of a “sticky trace” from self-binding: since information bound to the self-model is bound preferentially, it is harder to unbind later.

This empirically detectable signature of self-binding suggests the following prediction: Patients who associate neutral stimuli with self vs. others prior to receiving psilocybin or placebo, and are asked to re-associate that information after the drug session, should show performance differences on the latter task. Patients in the psilocybin condition should show less of a “sticky trace” from self-binding, if psilocybin disrupts this function. And if self-unbinding is the main *therapeutic* mechanism, the extent of this reduction in the “sticky trace” should correlate with symptom reduction in the subsequent days or weeks. This latter correlation would not be predicted by the RL hypotheses that Colombo sketches, according to which psychedelic therapy “mainly works not by changing self-representations, but by changing one’s representations of time or value” (Colombo, 2022). This is not to say that a self-binding explanation of a diminished “sticky trace” would be the only explanation possible. However, a strong correlation between a reduced “sticky trace” and good clinical outcomes would certainly fit better with a self-representation theory of psychedelic therapy than with a theory that denies that changes to self-representation are the main mechanism. Thus, without answering questions about the neural implementation of precision expectation and the computational function of serotonin, the self-representation hypothesis is specific enough to advance existing understanding of psychedelic therapy.

There is another point to make: even if psychedelics typically decrease the precision of self-representations via 5-HT_{2A}R agonism, this does not entail that serotonin or the 5-HT_{2A}R have any unique or distinctive involvement in self-representation or in precision estimation. It only entails that 5-HT_{2A}R hyperactivation is capable of inducing some set of neurocognitive effects that typically includes disruption to the substrates of self-representation and/or precision estimation. This is one way to understand the REBUS model and its antecedents, such as the entropic brain hypothesis (EBH; Carhart-Harris et al., 2014): as strictly agnostic about whether serotonin or the 5-HT_{2A}R have any special involvement in computational functions such as precision weighting. (Of course, proponents of these hypotheses do make some implementational speculations, but it is not clear that they intend these speculations to constitute core commitments of their accounts.) On the implementationally agnostic reading, the REBUS model would only be committed to the following: since the 5-HT_{2A}R is densely expressed in neural systems

that plausibly have the function of encoding high-level (self-related) priors, hyperactivating this receptor disrupts the functioning of these systems, leading *somehow* to those priors being assigned lower precision. There are multiple possible ways of filling in the blanks. The overall point is a familiar one in the philosophy of science: theoretical claims (such as the core claims of REBUS and EBH) must be conjoined with auxiliary (in this case, implementational) hypotheses in order to yield specific, testable predictions.

As Colombo notes, the plurality of suggestions in the PP literature about the computational function of serotonin and the neural implementation of precision estimation may plausibly “be taken to highlight that there is a plurality of pathways and mechanisms implementing precision-weighted computing in the brain” (Colombo, 2022). Equally plausibly, one may suppose that neuromodulators such as serotonin, and receptor subtypes such as the 5-HT_{2A}R, participate in multiple computational functions, such that (i) there is no one-to-one mapping between neuromodulator or receptor subtype and computational function, and (ii) understanding the effects of changes to neuromodulation requires attending to the plausible computational functions of relevant neural structures (Dietrich, 2003). These assumptions underlie the account I develop in POP. Of course, they may turn out to be false. Work should be done to formulate more specific and testable PP hypotheses about psychedelic therapy and pit these against alternative (e.g. RL) hypotheses. It is also worth considering whether the PP framework is genuinely incompatible with the RL hypotheses that Colombo outlines (Colombo, personal communication). But in any case, the pluralism in the PP literature about serotonin function and precision estimation implementation do not tell strongly against a PP account of psychedelic experience, unless we have reasons to rule out the plausible pluralistic possibility just sketched. Moreover, its abstraction from certain implementational details does not make the predictive self-binding account too abstract or generic to advance existing understanding of psychedelic therapy or to generate specific, testable predictions.

3 Epistemology

The commentaries that focus on epistemological issues also vary in how strongly they challenge my account; again, I discuss these commentaries in ascending order of the degree of challenge. Bortolotti and Murphy-Hollies endorse the overall shape of my account and make some productive connections to ideas concerning agency-first epistemology and self know-how. Caporuscio brings into play the idea of self-shaping, which connects to self know-how, but she argues against certain core claims of my account. Finally, Fink allows that psychedelic experiences can have some epistemic benefits, but argues that these experiences cannot contribute to the justification of beliefs, and that the epistemic benefit I call “new knowledge of old facts” is better conceptualized as *increased understanding*.

3.1 Agency first or agency also?

Bortolotti and Murphy-Hollies (B&M) (2022) endorse the broad outlines of my epistemology of psychedelics, and relate it to a recent trend towards an *agency-first* approach to epistemology. On such an approach we “move from asking whether the agent’s belief is true and well-grounded to asking whether the belief enables the agent to succeed in their epistemic practices” (Bortolotti & Murphy-Hollies, 2022). As B&M say, this mode of epistemic evaluation is one that I deploy in POP and is necessary to understand the overall epistemic profile of psychedelic experiences.

It strikes me as apt and illuminating to relate aspects of my methodology to these broader currents in epistemology. My main disagreement with B&M is a relatively minor point of emphasis and priority: I do not think all the epistemic benefits of psychedelics I postulate in the book can best, or only, be captured by an agential approach. Rather than an agency-first approach, I favour an *agency-also* approach that sees agential and more traditional forms of epistemic evaluation as equally important.⁴ Indeed, I think the value of agential evaluation derives from the value of more traditional evaluation: it is *because* we care about justification, truth, reliability etc. that we should want to evaluate, not just the justification, truth, etc. of individual cognitions, but also the *overall effects* of those cognitions on the individual’s propensity to form or acquire cognitions that are justified, true, etc. The fundamental values here are still those of traditional epistemology; we are simply adopting a broader perspective.

In chapter 8 of POP I argue that therapeutic psychedelic experiences often have five distinct types of epistemic benefits: the acquisition of important forms of knowledge that (propositional knowledge), knowledge how (ability knowledge), knowledge by acquaintance, and new knowledge of old facts, as well as indirect epistemic benefits. I think types 1 and 5 – propositional knowledge and indirect epistemic benefits – are best captured by an agential approach. The reason why is obvious in the case of type 5, indirect benefits: these precisely have to do with the effects of psychedelic experiences on individuals’ epistemic agency.

Type 2, knowledge how, probably has beneficial effects on epistemic agency, but it is also epistemically valuable in its own right, and this can be accommodated comfortably by a traditional approach. It is true that traditional epistemology focuses disproportionately on propositional knowledge, but knowledge how is standardly regarded as a form of knowledge in its own right. Similar remarks apply to types 3 and 4, knowledge by acquaintance and new knowledge of old facts. The psychedelic-induced acquisition of these plausibly has beneficial effects

⁴Of course, there is more than one way of understanding what an agency-first approach amounts to. It might be a matter of causal or explanatory priority, suggesting that being a successful epistemic agent is necessary to acquire traditional epistemic goods, or it might be a matter of normative or evaluative priority, suggesting that epistemic evaluation ought to focus more strongly on overall epistemic agency than on the truth, justification, etc. of specific mental states. (I am indebted to Chiara Caporuscio for raising this point.) My reading of B&M is in line with the latter construal, and this is reflected in my discussion.

on epistemic agency – but those indirect effects belong in category five. The intrinsic epistemic value of knowledge by acquaintance and new knowledge of old facts can be captured in traditional epistemological terms, independently of their agential effects.

Why think that type 1, propositional knowledge, fits well with an agential approach? This has to do with the ideas that B&M introduce about self know-how (Bruin & Strijbos, 2020). I argue that in some cases psychedelic experience, plus subsequent reflection, can help people to gain new justified true beliefs about their own psychological states and traits. This increases the coherence between their models of themselves and the reality that those models represent (i.e. the person herself and her patterns of thought, feeling, and behaviour). But as B&M, following DeBruin and Strijbos, point out, there is another way that we can increase the coherence between self-models and reality: shaping ourselves in accordance with those models. I think it is plausible that psychedelic experiences can improve this kind of *self know-how*, and that this epistemic benefit is best captured by an agential lens. These ideas connect with Caporuscio's arguments about self-shaping.

3.2 Self shaping and self know-how

B&M, following DeBruin and Strijbos (2020), note that one way we can preserve the authority of first-person ascriptions is by making our behaviour line up with them: a world-to-mind, rather than a mind-to-world, direction of fit. This brings to mind the notion of active inference from the PP literature. It also provides a helpful way to think about certain psychedelic phenomena. Psychedelic subjects often have apparent insights into the nature of their *true* or *authentic* selves, which raises difficult questions, especially when the content of these insights may be at odds with subjects' recent patterns of thought, feeling, and behaviour. But even when such insights are strictly false, one way to make them true is to act in accordance with them, so that they become self-fulfilling prophecies. Perhaps the truth in such insights is not that this is *who I really am*, but that this is *someone I can be*.

Caporuscio's (2022) motivation to invoke ideas of this kind comes from a critique of my epistemological claims. I argue that, by decreasing the precision of self-related priors, psychedelics probably sometimes lead to accurate insights concerning the self. Caporuscio asks: why should we think psychedelic insights concerning the self are any more likely to be accurate than those concerning the external world? After all, if the REBUS model is true, then high-level self-related and world-related priors get relaxed equally. Thus, if we treat world-related insights as suspect, then why not treat self-related insights the same way?

I think there are some reasons to think that self-related insights are likelier to be accurate than (external) world-related insights. For one thing, we know that many of psychedelics' effects on external-world perception tend in a misrepresentational direction, such as the perception of motion in static objects and geometrical patterns overlaying objects. So there are positive grounds for suspicion in

the external world case that are not present to the same degree in the introspective case. (Of course, we know that many self-related beliefs are inaccurate, but this is not the same as knowing that psychedelics make them more likely to be inaccurate.) Second, it seems plausible to me that some of psychedelics' effects on introspection simply amount to *promoting* it, rather than qualitatively altering it. In this light, we ought to remember that introspection, in general, *can* provide defensible justification for self-related beliefs, and being on a drug is not, in itself, an automatic defeater (Letheby, 2021, p. 163). Third, insofar as psychedelics do qualitatively alter introspection, the alterations they induce are ones that seem likely to improve its accuracy. It is plausible that self-related affective biases, of the kind that psychedelics can disrupt via their effects on self-modelling, ordinarily impede self-knowledge (Davis & Thompson, 2015); thus, psychedelics may reduce obstacles to self-knowledge. (Of course, self-related priors, like external world-related priors, are epistemically enabling as well as constraining; thus, relaxing them is sure to have some epistemic costs, at least for the duration of the experience.)

In POP I suggest that relaxing self-related priors, such as negative *core beliefs*, can allow evidence inconsistent with those priors into consciousness. Caporuscio agrees that this can happen, but suggests that it is not the whole story. In some cases, she says, there may be little evidence inconsistent with the relevant priors if we have acted in accordance with them for a long time. Rather, relaxing these priors allows us to explore *new* self-conceptions; we can then shape ourselves in accordance with them during the integration period.

I think Caporuscio is right to draw attention to this kind of case, but I also think many cases of this kind can be accommodated by my account, if we understand the notion of "evidence" sufficiently broadly. Take Caporuscio's case, adapted from Moran (2001) and McGeer (1996), in which Maria has conceived of herself as self-sufficient and emotionally independent for all of her adult life, and has acted, thought, and felt in accordance with that self-conception, so that behavioural and psychological evidence against it is scant. On psilocybin, she is released from the constraint of this self-conception and explores new ways of being, seemingly discovering a deep, heretofore unknown desire for human connection. One possibility is that the desire was there all along, but was repressed, unattended, or otherwise unknown. I am not sure what sort of psychodynamic posits this possibility requires, but any plausible model of the mind must account for the fact that we often have desires of which we are, in some sense, unaware. Another possibility is that the desire did not predate Maria's psilocybin session, but was created anew by it. Still, merely by responding positively to the thought of closer human connection, Maria gains some information (i.e. evidence) about her own dispositions and capacities – evidence that her ordinary self-model would have prevented her from gaining.

The objection might still be raised that Maria would be unjustified in concluding that the desire had always been there, i.e., that it antedated the psilocybin session. I agree: without further evidence, she would – though she may well dis-

cover such evidence as she reflects on her past in a new light. In any case, Caporuscio and B&M are correct in this much: Sometimes, what people discover on psychedelics is not facts about *who they really are* underneath layers of avoidance and self-deception, but facts about *who they can be* and *who they want to be*. However, these are modal and psychological facts for which one can discover *evidence*, in the form of the availability or emergence of certain patterns of thinking and feeling, when the everyday priors that would normally inhibit those patterns are relaxed.

Again, too, it is worth noting that many patients do not discover *new* forms of self-modelling, but instead, old, disused ones: “At its most basic, I feel like I did before the depression.” (Watts et al., 2017). Patients often describe remembering, or reconnecting with, old passions, pursuits, and hobbies. Cases like these seem straightforwardly epistemically beneficial, and allow little room for scepticism: people are not discovering some new, hidden fact about themselves, but *being reminded of* a fact that they had forgotten, in the banal sense that they had not thought about it for a long time. This is not the same as discovering apparently repressed childhood memories: in the present kind of case, there is usually no reason to distrust the memory. Of course, it *could* have occurred without psychedelics – but whether it *would* have is unclear. This is a genuine epistemic benefit – simply being reminded of a known, important, but long-forgotten or long-unnoticed fact⁵.

Caporuscio argues that the idea of self-shaping casts the epistemic and psychological importance of the integration period in a new light: this period is not merely about scanning putative insights for plausibility, but also about determining which (re-) discovered self-conceptions one wants to commit to. I agree, and Caporuscio is right to emphasise this. The benefits of psychedelics for self-knowledge include both (i) discovering facts about who one is, and (ii) discovering facts about who one *can* be, then shaping oneself accordingly. However, we should note that often, choosing which new self-conceptions to adopt may not be a conscious, deliberate process. As L&F point out, such insights can carry a noetic quality – people *feel* as though they have discovered who they really are. And in a cognitive dynamic that justifies the appellation “inverse PTSD” (Garcia-Romeu, R. Griffiths, & W. Johnson, 2014), the salience and positive valence of this experience compel subjects to revisit it, causing lasting and significant changes to their thoughts, feelings, and behaviour. Self-shaping happens during integration, but it need not always be a conscious or deliberate act.

3.3 Understanding and justification

Fink (2022) criticizes some of my views concerning psychedelic epistemology, and advances a positive proposal of his own. His positive proposal is that the epistemic benefit of psychedelic experiences that I call “new knowledge of old facts” can

⁵I am indebted to Matthew Nestor for bringing this sort of case to my attention.

better be conceptualized as *increased understanding*. This, he notes, has the virtue of generality, covering cases in which the proposition in question is false. I am sympathetic to this line of thought; indeed, I have expressed this sympathy briefly in print before (Letheby, 2019). I think that when the proposition in question is true, it is also illuminating to speak of “new knowledge of old facts” – but I will not defend that claim here. Instead, I will focus mainly on Fink’s arguments that psychedelic experiences cannot contribute to the justification of beliefs.

First, however, I will make one other point briefly: Fink holds that increased understanding is an *indirect* epistemic benefit, not a direct one – but this depends on how one defines “direct” and “indirect”. In POP my use of these terms was intended to draw a distinction between (a) epistemic benefits that result from the psychedelic experience itself and (b) epistemic benefits that result, instead, from the lasting psychological *effects* of the psychedelic experience. On this way of drawing the direct/indirect distinction, increased understanding is, indeed, a direct epistemic benefit of psychedelic experiences. (Fink draws this distinction differently; I will return to this point shortly.)

What, then, of Fink’s arguments concerning justification? To begin with, he agrees with my claim that psychedelics can enhance the context (or processes) of discovery – in other words, even if they cannot give us justification for beliefs, they can help us come up with promising new ideas. However, since there are many processes of discovery, he says, this benefit “cannot be considered unique to psychedelics” (Fink, 2022). Here, I want to sound a note of caution. It is true that people can come up with promising new ideas in many different ways. But this is consistent with the idea that some of these processes are more powerful or effective than others. And I strongly suspect that psychedelic ingestion may be more effective at helping people to come up with radically new, creative, or interesting ideas than many other methods. This, of course, is an empirical claim, and a body of research already exists examining psychedelics’ effects on creativity (Girn, Mills, Roseman, Carhart-Harris, & Christoff, 2020; Sessa, 2008) – but examining this literature is beyond the scope of this paper. Nonetheless, it seems plausible that many people who take psychedelics, especially in deliberate, intentional ways for therapeutic, transformative, or problem-solving purposes, thereby come up with important and valuable ideas that they would not have come up with otherwise.

Fink also argues that, because psychedelic ingestion can lead to the acquisition of false beliefs as well as knowledge, its contribution – even in cases in which knowledge results – may not be *epistemic*: “because some of the processes altered by psychedelic compounds that lead us to novel mental states may also lead us to comfortable illusions or placebo insights, I would not consider this contribution necessarily epistemic. The contribution of psychedelics to knowledge, if there is any, is coincidental” (Fink, 2022). This argument strikes me as invalid. Every method of knowledge acquisition – consider, for example, silent mathematical reasoning – is fallible, and capable of generating false beliefs as well as knowledge. But surely it would be too quick to conclude, *on this basis*, that none of these methods make genuinely epistemic contributions. Similar points apply when we

compare psychedelic ingestion to other processes of discovery. If one engages in some process of discovery in order to come up with a new idea – say, an insight into one’s personality or a solution to some problem – and then comes up with an idea, subsequently tests it, acquires justification, and ends up with new propositional knowledge, this epistemic outcome (i.e. the acquisition of propositional knowledge) can hardly be viewed as a coincidence. True, processes of discovery must be combined with processes of justification; but this does not entail that the former can only make coincidental contributions to our epistemic state.

Fink’s central critical claim about psychedelic epistemology is the thesis of *Psychedelic Justification Impossibilism* (PJI):

Psychedelic Justification Impossibilism: “psychedelics may have psychological benefits, indirect epistemic benefits, or be beneficial for processes of discovery... but they lack direct epistemic benefits for processes of justification” (2022)

Fink defines “direct” epistemic benefits as those which contribute to a mental state’s status as knowledge. In the case of propositional knowledge, he understands this contribution in purely justificational terms: he says that the psychedelic justification impossibilist may

grant indirect epistemic benefits in that psychedelic compounds, experiences, or episodes may contribute to forming states that, as a matter of fact, are knowledge but deny direct epistemic benefits *in that they contribute to such states’ status as knowledge, i.e. to processes of justification* (Fink, 2022)

So, in essence, PJI is the claim that psychedelic ingestion and psychedelic experiences cannot contribute to the epistemic justification that is required for a true belief to constitute propositional knowledge.

Before examining Fink’s arguments for PJI, it is worth noting that the thesis is consistent with most of my epistemological claims about psychedelics. Fink suggests that the points he makes may generalize to other types of knowledge besides the propositional. But four of the five types of epistemic benefits I attribute to psychedelic experiences – knowledge how, knowledge by acquaintance, new knowledge of old facts, and indirect epistemic benefits – are ones which are not standardly conceived of as requiring *justification*. And when it comes to propositional knowledge – the type about which I am most tentative and ambivalent in the book – my most confident claim is that psychedelics have substantial benefits for discovery. As Fink notes, this claim would be untouched by the truth of PJI. (And, as indicated above, I think benefits for discovery are genuinely epistemic.) So even if PJI is true, this does not undermine the idea that therapeutic and transformative psychedelic experiences typically have significant epistemic benefits that are consistent with a naturalistic worldview.

Fink contests the idea that PJI will not generalize to other types of knowledge. Justification, he argues, can be needed for some other forms of knowledge as well, “e.g., when we need to justify that the success of an act [...] is not just lucky but due to know-how or skill” (Fink, 2022). But it seems to me that the kind of justification at issue here is not justification *for the knowledge how*; it is justification for *the claim that someone possesses knowledge how* – i.e. justification for a proposition. The fact that epistemic justification is needed for our (propositional) attributions of non-propositional types of knowledge does not show that justification is needed to possess those types of knowledge.

All this aside, the truth or falsity of PJI is clearly a matter of considerable interest and importance. While I suggest in POP that psychedelic experiences alone rarely, if ever, provide *sufficient* justification for interesting self-related beliefs, I argue that they can make a substantial justificatory contribution. Thus, there is a real point of disagreement here.

What, then, of Fink’s arguments for PJI? One argument appeals to the idea that standards of justification are fixed extra-individually, and therefore cannot be affected by an individual’s act of psychedelic ingestion:

[Epistemic] justification has a social and normative dimension [...] what amounts to a good or adequate justification [is] largely independent of any given individual because the standards for justification are social or extra individual. And because psychedelics affect only an individual’s state, leaving the standards of justification untouched, they do not affect which processes can be counted as processes of justification even if processes of justification are in the end psychological processes and indeed affected by psychedelics. The epistemic lifting is done not by psychedelically influenced processes, but by the extra individual circumstances that turn these processes into processes of justification. (Fink, 2022).

This argument about affecting standards of justification seems to me to be a red herring. Of course psychedelic ingestion does not affect which processes can be counted as processes of justification – but this is surely not what any sensible proponent of psychedelic epistemology has in mind. None of our other knowledge acquisition methods – sense perception, reasoning, testimony, etc. – benefit individuals epistemically by affecting which processes *can be counted as* processes of justification. Rather, they benefit individuals epistemically by (*inter alia*) causing them to engage in processes that are, in fact, processes of justification. When I proposed in POP that psychedelic ingestion contributes somewhat to the context of justification, I did not mean that psychedelic ingestion *alters the standards* of epistemic justification. Rather, I meant that psychedelic ingestion, by affecting individuals’ psychological states, can help those individuals to *satisfy* the existing standards of epistemic justification.

Fink acknowledges that processes of justification may be psychological processes that could turn out to be affected – perhaps for the better – by psychedelics; surely this is as much as we could ask, vis-à-vis justification, from any method of knowledge acquisition. He claims that the “epistemic lifting is done not by psychedelically influenced processes, but by the extra individual circumstances that turn these processes into processes of justification” (Fink, 2022) – but this sort of claim seems to me implausible when applied to other epistemic practices. Consider, again, silent mathematical reasoning. Surely this can do some fairly heavy epistemic (justificatory) lifting, even if it does not alter the *standards* of justification one whit.

Fink mentions reliabilism as one position in epistemology that might licence the claim that psychedelics contribute to processes of justification. Here, I agree. Reliabilism broadly holds that propositional knowledge is true belief formed by a reliable process. Part of my strategy in POP was to appeal to certain facts about the psychedelic experience and the integration period to argue that these two, *combined*, can be a fairly reliable way of attaining accurate insights about the self. Fink disagrees, arguing that “the high prevalence of placebo insights and comforting delusions [...] makes it implausible that psychedelically altered mechanisms are reliable” (Fink, 2022). But when it comes to putative insights concerning the self, we do not know what the prevalence of placebo insights relative to genuine insights is – this is precisely the question at issue. I argue (tentatively) on various grounds that the ratio of genuine self-related insights to placebo self-related insights is likely to be fairly high – perhaps even high enough to pass some relevant reliability threshold for epistemic justification. I do not think we can embrace this conclusion confidently, but nor can we embrace its negation any more confidently. The frequency with which controlled psychedelic ingestion promotes accurate vs. inaccurate psychological or self-related insights is an important matter for future research, both philosophical and empirical. Determining this frequency poses formidable methodological difficulties; if any progress can be made, it is likely to require concerted collaborative efforts from experts in multiple disciplines (cf. Tulver, Kaup, Laukkonen, & Aru, 2021).

Finally, I think it is important to consider what exactly PJI means, and what it would take to falsify it. Suppose, as I suggested in subsection 3.2 above, that one of psychedelics’ main effects on introspection is simply to promote it: i.e. to induce a more introspective state of consciousness than is usual. Suppose that you ingest psilocybin, become temporarily more introspective, and thereby acquire a new true belief about yourself that you would not have acquired otherwise. Now, introspection in general can provide defeasible justification for self-related beliefs, and I argued in POP that being on a drug is not *in itself* an automatic defeater. What, then, are we to say about such a case? It seems clear to me that, in a case like this, you have a justified⁶ true belief such that you would not have formed the belief, *nor* obtained the justification, had you not taken psychedelics. (Similar

⁶Unless some other defeater is present.

points can be made regarding cases wherein most, or all, of the justification derives from sober reasoning after the experience.) Is it not straightforwardly true in such a case that psychedelic use benefited you epistemically – not just by causing you to acquire a true belief, but by causing you to acquire a *justified* true belief? According to counterfactual theories of causation, at least, your psychedelic use certainly caused you to acquire these epistemic benefits in such a case.

In sum, I agree with Fink that the epistemic benefit of psychedelics that I call “new knowledge of old facts” can be conceptualized at least as usefully, and with greater generality, in terms of increased understanding. I also agree that, while psychedelics’ benefits for epistemic discovery are fairly clear, it is less clear what contribution, if any, they can make to justification. I argued in POP that they can make a substantive contribution, whereas Fink argues that they cannot – this, of course is the thesis of PJI. I think it is fair to say that neither of our arguments settles the issue decisively. As such, I conclude by hazarding one final point of agreement: that further research on this intriguing question is required.

4 Conclusion

All commentators raise important and interesting questions about psychedelics and about the specific account developed in POP. Do we need to pay more attention to the roles of affect and strengthened priors to understand psychedelic therapy? Is phenomenal opacity more important than transparency in psychedelic therapy? Is PP the right framework to explain psychedelic therapy, and how do computational hypotheses relate to neural implementation details? Does psychedelic epistemology require an agency-first approach? What roles might self-shaping and self know-how play in psychedelic epistemology? Does psychedelic experience provide epistemic justification or increased understanding?

Here I have offered some initial thoughts on these questions. On the whole, this symposium opens up many intriguing possibilities for the philosophy (and science) of psychedelics. I hope that it stimulates more discussion on these matters both within and across disciplines. I would like to finish by once again thanking all the commentators for their generous contributions.

Acknowledgments

For helpful comments and suggestions on earlier versions, I would like to thank Chiara Caporuscio, Matteo Colombo, Sascha Fink, and Sarah Hoffman. I would also like to thank Chiara Caporuscio, Sascha Fink, and everyone else at *Philosophy and the Mind Sciences* for generously volunteering to edit/host this symposium and working tirelessly to make it a reality. This research was partially supported by the Australian Government through the Australian Research Council’s Discovery Projects funding scheme (project DP190101451). The views expressed herein are those of the author and are not necessarily those of the Australian Government or Australian Research Council.

References

- Bortolotti, L., & Murphy-Hollies, K. (2022). The agency-first epistemology of psychedelics. *Philosophy and the Mind Sciences*, 3. <https://doi.org/10.33735/phimisci.2022.9283>
- Bruin, L. de, & Strijbos, D. (2020). Does confabulation pose a threat to first-person authority? Mindshaping, self-regulation and the importance of self-know-how. *Topoi*, 39(1), 151–161. <https://doi.org/10.1007/s11245-019-09631-y>
- Caporuscio, C. (2022). Belief now, true belief later: The epistemic advantage of self-related insights in psychedelic-assisted therapy. *Philosophy and the Mind Sciences*, 3. <https://doi.org/10.33735/phimisci.2022.9310>
- Carhart-Harris, R., Leech, R., Hellyer, P., Shanahan, M., Feilding, A., Tagliazucchi, E., et al. (2014). The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00020>
- Colombo, M. (2022). Serotonin, predictive processing and psychedelics. *Philosophy and the Mind Sciences*, 3. <https://doi.org/10.33735/phimisci.2022.9320>
- Davis, J. H., & Thompson, E. (2015). Developing attention and decreasing affective bias: Toward a cross-cultural cognitive science of mindfulness. In *Handbook of mindfulness: Theory, research, and practice* (pp. 42–61). New York, NY, US: The Guilford Press.
- Dietrich, A. (2003). Functional neuroanatomy of altered states of consciousness: The transient hypofrontality hypothesis. *Consciousness and Cognition*, 12(2), 231–256. [https://doi.org/10.1016/S1053-8100\(02\)00046-6](https://doi.org/10.1016/S1053-8100(02)00046-6)
- Fink, S. B. (2022). Psychedelics favour understanding but not knowledge. *Philosophy and the Mind Sciences*, 3. <https://doi.org/10.33735/phimisci.2022.9320>
- Gallimore, A. R., & Strassman, R. J. (2016). A model for the application of target-controlled intravenous infusion for a prolonged immersive DMT psychedelic experience. *Frontiers in Pharmacology*, 7. <https://doi.org/10.3389/fphar.2016.02111>
- Garcia-Romeu, A., R. Griffiths, R., & W. Johnson, M. (2014). Psilocybin-occasioned mystical experiences in the treatment of tobacco addiction. *Current Drug Abuse Reviews*, 7(3), 157–164. <https://doi.org/10.2174/1874473708666150107121331>
- Girn, M., Mills, C., Roseman, L., Carhart-Harris, R. L., & Christoff, K. (2020). Updating the dynamic framework of thought: Creativity and psychedelics. *NeuroImage*, 213, 116726. <https://doi.org/10.1016/j.neuroimage.2020.116726>
- Hendricks, P. S. (2018). Awe: A putative mechanism underlying the effects of classic psychedelic-assisted psychotherapy. *International Review of Psychiatry*, 30(4), 331–342. <https://doi.org/10.1080/09540261.2018.1474185>
- Hoffman, S. (2022). Positive affect and Letheby's naturalization of psychedelic therapy. *Philosophy and the Mind Sciences*, 3. <https://doi.org/10.33735/phimisci.2022.9285>
- Johnson, M. W. (2022). Psychedelic science needs philosophy. *Philosophy and the Mind Sciences*, 3. <https://doi.org/10.33735/phimisci.2022.9452>
- Letheby, C. (2019). The Varieties of psychedelic epistemology. In N. Wyrđ, D. Luke, A. Tollan, C. Adams, & D. King (Eds.), *Psychedelics: More food for thought from Breaking Convention*.
- Letheby, C. (2021). *Philosophy of Psychedelics*. Oxford University Press.
- Lyon, A., & Farennikova, A. (2022). Through the psychedelic looking-glass: The importance of phenomenal transparency in psychedelic transformation. *Philosophy and the Mind Sciences*, 3. <https://doi.org/10.33735/phimisci.2022.9323>
- Martin, J. M., & Sterzer, P. (2022). How level is the 'Cognitive Playing Field'? Context shapes alterations in self-conception during the psychedelic experience. *Philosophy and the Mind Sciences*, 3.
- McGeer, V. (1996). Is "self-knowledge" an empirical problem? Renegotiating the space of philosophical explanation. *The Journal of Philosophy*, 93(10), 483–515. <https://doi.org/10.2307/2940837>
- Metzinger, T. (2004). *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.
- Moran, R. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. <https://doi.org/10.1515/9781400842971>
- Nichols, D., Johnson, M., & Nichols, C. (2017). Psychedelics as medicines: An emerging new paradigm. *Clinical Pharmacology & Therapeutics*, 101(2), 209–219. <https://doi.org/10.1002/cpt.557>
- Noorani, T., Garcia-Romeu, A., Swift, T. C., Griffiths, R. R., & Johnson, M. W. (2018). Psychedelic therapy for smoking cessation: Qualitative analysis of participant accounts. *Journal of Psychopharmacology*, 32(7), 756–769. <https://doi.org/10.1177/0269881118780612>
- Revonsuo, A. (2006). *Inner Presence: Consciousness as a Biological Phenomenon*. Cambridge, MA, US: MIT Press.
- Safron, A. (2020). On the varieties of conscious experiences: Altered beliefs under psychedelics (ALBUS). *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/zqh4b>
- Sessa, B. (2008). Is it time to revisit the role of psychedelic drugs in enhancing human creativity? *Journal of Psychopharmacology*, 22(8), 821–827. <https://doi.org/10.1177/0269881108091597>

Letheby, C. (2022). Self and Knowledge in Psychedelic Therapy: Reply to Commentaries on *Philosophy of Psychedelics*. *Philosophy and the Mind Sciences*, 3, 12.

<https://doi.org/10.33735/phimisci.2022.9642>



©The author(s). <https://philosophymindscience.org> ISSN: 2699-0369

- Sui, J., & Humphreys, G. W. (2015). The integrative self: How self-reference integrates perception and memory. *Trends in Cognitive Sciences*, 19(12), 719–728. <https://doi.org/10.1016/j.tics.2015.08.015>
- Tulver, K., Kaup, K. K., Laukkonen, R., & Aru, J. (2021). Restructuring insight: An integrative review of insight in problem-solving, meditation, psychotherapy, delusions and psychedelics. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/8fbt9>
- Wang, H., Humphreys, G., & Sui, J. (2016). Expanding and retracting from the self: Gains and costs in switching self-associations. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2), 247. <https://doi.org/10.1037/xhp0000125>
- Watts, R., Day, C., Krzanowski, J., Nutt, D., & Carhart-Harris, R. (2017). Patients' accounts of increased "connectedness" and "acceptance" after psilocybin for treatment-resistant depression. *Journal of Humanistic Psychology*, 57(5), 520–564. <https://doi.org/10.1177/0022167817709585>

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Letheby, C. (2022). Self and Knowledge in Psychedelic Therapy: Reply to Commentaries on *Philosophy of Psychedelics*. *Philosophy and the Mind Sciences*, 3, 12.

<https://doi.org/10.33735/phimisci.2022.9642>



©The author(s). <https://philosophymindscience.org> ISSN: 2699-0369