PhiMiSci
Philosophy and the Mind Sciences

Serotonin, Predictive Processing, and Psychedelics

Matteo Colombo[a] (m.colombo@uvt.nl)

Abstract
Letheby's Philosophy of Psychedelics relies on Predictive Processing to try and find unifying explanations relevant to understanding how serotonergic psychedelics work in psychiatric therapy, what subjective experiences are associated with their use and whether such experiences are epistemically defective. But if Predictive Processing lacks genuinely explanatory unifying power, Letheby's account of psychedelic therapy risks being unwarranted. In this commentary, I motivate this worry and sketch an alternative interpretation of psychedelic therapy within the Reinforcement Learning framework.

Keywords
Predictive Processing · Reinforcement Learning · Reinforcement Learning · Unification

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.

Chris Letheby's (2021b) *Philosophy of Psychedelics* relies on Predictive Processing (PP) to try to find unifying explanations that could bridge computational function, neural mechanism and phenomenological analysis relevant to understanding how serotonergic psychedelics work in psychiatric therapy, which subjective experiences are associated with their use and whether these experiences are epistemically defective. But if PP lacks genuinely explanatory (counterfactual supporting) unifying power, then Letheby's account of psychedelics will rest on fragile bridges between computation, mechanism and phenomenology; and with such fragile bridges, one might worry that Letheby's interpretation of psychedelic therapy in terms of PP is unwarranted. In this commentary, I motivate this worry and sketch an alternative interpretation of psychedelic therapy within a computational framework different from PP.

[a]Tilburg University

Colombo, M. (2022). Serotonin, Predictive Processing, and Psychedelics. *Philosophy and the Mind Sciences*, 3, 7. https://doi.org/10.33735/phimisci.2022.9320

Virtually all researchers working within the PP framework agree that some psychiatric conditions originate in and are maintained by the malfunctioning of precision-weighted prediction error signalling (Colombo & Fabry, 2021; see e.g., Corlett et al., 2009). This disruption would result in too much (or too little) weight being given to sensory evidence compared with prior "beliefs"; and due to this misallocation, people would form delusional beliefs that could explain away overly salient perceptions (or would tend to perceive only what they are expecting to perceive).

In particular, PP models such as Carhart-Harris & Friston's (2019) "RElaxed Beliefs Under pSychedelics" and Letheby & Gerrans's (2017) self-binding model of psychedelic ego dissolution share the idea that serotonergic psychedelics disrupt "predictive models implemented in high-level cortical networks… and this disruption creates an opportunity to revise these models in beneficial ways" (Letheby, 2021a, p. 9). For Letheby, psychedelics weaken the precision of maladaptive self-representations (i.e., mental representations of one's self), which are often associated with anxiety disorders, depression and addiction; and this weakening would provide psychiatrists with an opportunity for therapeutic interventions aimed at "changing how patients see themselves and relate to their own minds and lives" (Letheby, 2021a, p. 12).

One set of questions relevant to evaluate the empirical adequacy of this suggestion, as well as the extent to which PP provides us with a genuinely unifying explanatory bridge for understanding the workings of psychedelic therapy is this: What's the function of serotonin, which is the primary target of psychedelics? What kind of information does serotonergic neuromodulation carry? Does serotonin implement precision estimates, precision-weighted high-level priors or some other aspects of precision-weighted computing?

Researchers relying on PP have made different suggestions about these questions. Some posit that the computation of precision estimates is implemented by the activity of dopaminergic neurons (FitzGerald et al., 2015; Friston et al., 2014). Others suggest that increased acetylcholine activity enhances sensory precision by amplifying bottom-up signals when external stimuli are predictable (Moran et al., 2013); other researchers that both noradrenaline and dopamine are implicated in implementing estimates of the precision of sensory evidence and in tracking the volatility of the environment (Yon & Frith, 2021); and others yet hypothesise that precision estimates are computed and manipulated by the orchestrated activity of dopamine, acetylcholine and serotonin neurons (Adams et al., 2013; Clark, 2013).

These different suggestions within the PP literature could be taken to indicate that computational posits like *precision-weighted high-level priors* are somehow ill-defined. But this conclusion is too quick, as the notions of *precision* and *high-level prior* are precisely and explicitly defined within many particular PP models (aka active inference models) of particular psychiatric phenomena (see e.g., Lawson et al., 2017).

More plausibly, the plurality of suggestions within the PP literature could be taken to highlight that there is a plurality of pathways and mechanisms imple-

Colombo, M. (2022). Serotonin, Predictive Processing, and Psychedelics. *Philosophy and the Mind Sciences*, 3, 7. https://doi.org/10.33735/phimisci.2022.9320

menting precision-weighted computing in the brain. This plurality of pathways and mechanisms for precision-weighted computing is consistent with the evidence from the neuroimaging studies that Letheby (2021b, Chapter 5) and Letheby & Gerrans (2017) review, which indicates the existence of an association between psychedelic therapy and changes in the *Default Mode* and *Salience networks* in the brain. It is consistent also with the *Molecular Neuroplasticity Theory* (Letheby, 2021b, Chapter 4) discusses, whereby serotonin and particularly 5-HT2A receptors (one among a dozen known subtypes of serotonin receptors) have many distinct effects at several spatial and temporal scales—for example, serotonin activity influences both excitatory and inhibitory neurotransmission in neural circuits in the forebrain and midbrain, and modulates the release of other neurotransmitters like dopamine and acetylcholine, which impact neural plasticity and play various roles in learning (Hoyer et al., 2002; Jacobs & Fornal, 1999).

While precision-weighted high-level priors might be implemented across different structures of the nervous system, the current lack of knowledge of how they are actually implemented should not suggest that a computational model positing precision-weighted representations is untestable or that serotonin cannot contribute to implement those representations. But the lack of consistency within the PP literature about the computational function of serotonin and the putative physical realizers of precision estimates and precision-weighted high-level priors could also indicate that the bridge between computation, mechanism and phenomenology offered by PP is gappy and with limited explanatory power, which would call into question Letheby's choice to use PP to work out a unified, explanatory account of psychedelic therapy.

The worry here is that PP offers no more than re-descriptions of mental and neural phenomena. While these re-descriptions might be consistent with known empirical data, they would fall short as explanatory accounts that are experimentally tested and empirically evaluated against alternative hypotheses grounded in different theoretical frameworks (Colombo & Wright, 2017; Litwin & Miłkowski, 2020).

One such alternative framework is Reinforcement Learning (RL), which is increasingly popular in psychiatry (Colombo & Heinz, 2019; Maia & Frank, 2011). Compared to PP models, RL models have been submitted to experimental testing more frequently and more stringently, and they have proved themselves to enjoy a substantial degree of genuinely explanatory unifying power in the sciences of mind and brain (Colombo, 2014). Like PP models, any RL model of a mental phenomenon is incomplete, abstracts away from mechanistic details and its supporting evidence is mixed. Unlike PP models, however, RL modelling is grounded in the notions of *value* and *reward*, which PP models eliminate reinterpreting planning and decision-making as forms of inference (Colombo, 2016).

Now, the posits of *value* and *reward* in some RL models have been mapped onto various aspects of serotonin activity in algorithmic hypotheses about how the brain generates affective dynamics that drive learning and decision making

Colombo, M. (2022). Serotonin, Predictive Processing, and Psychedelics. *Philosophy and the Mind Sciences*, 3, 7. https://doi.org/10.33735/phimisci.2022.9320

(Dayan, 2012; Doya, 2002). Specifically, according to one of these hypotheses, serotonin represents expectations about negative outcomes (i.e., punishments or negative rewards) and, on the basis of these representations, it would compute appropriate inhibitory motor responses (Dayan & Huys, 2009). If this hypothesis is true, increased serotonin activity will promote the downregulation of impulsive emotional reactions to expected punishment or provocation, boosting patience and coping with anxiety. While this hypothesis has received some empirical corroboration (Crockett et al., 2009), more recent theoretical and empirical work suggests that serotonin plays more complicated functions, some in opposition and some in synergy with the functions played by dopamine in learning and decision making, depending on the agent's expectations about the statistical structure and controllability of the environment (Boureau & Dayan, 2011; Liu et al., 2020).

According to another hypothesis developed within the RL framework, serotonin computes "the time and resources available for action, learning and development" (Doya et al., 2021, p. 121). If this hypothesis is true, then serotonin activity is involved in computing multi-dimensional representations of the temporal horizon and the value of different outcomes at different time scales in different cognitive and biological tasks. Higher serotonin activity in a decision-making task would signal that there is enough time available for deeper deliberation, more explorative choices and longer-term predictions of the likely outcomes of different choices (Doya et al., 2021, Table 1).

Both RL hypotheses just sketched dovetail computational function, algorithmic processes and neural mechanism, while they also integrate affective and temporal dimensions that are central to the phenomenology of psychedelic experiences (Letheby, 2021b, Chapter 3). If we exclude the ecumenical possibility that separate networks of serotonin neurons and different subtypes of serotonin receptors perform different kinds of computational functions, then those RL hypotheses are not obviously consistent with the hypothesis that serotonin implements precision estimates and is involved in the weakening of the precision of high-level priors. But how would these hypotheses, if true, reconfigure Letheby's key idea that "psychedelic therapy works mainly by changing self-representation" (Letheby, 2021b, p. 6)?

One possible answer is that the truth of either of these RL hypotheses would *not* change anything in Letheby's key idea, because time, value and affect are core aspects of self-representations and self-representations might be implemented in high-level cortical networks encoding "the most fundamental assumptions concerning self, [but also] space, time, causality, and so forth: the basic parameters of our phenomenal worlds" (Letheby, 2021a, p. 8). 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.

A different answer is that the truth of either of these RL hypotheses indicates that psychedelic therapy mainly works *not* by changing self-representations, but

Colombo, M. (2022). Serotonin, Predictive Processing, and Psychedelics. *Philosophy and the Mind Sciences*, 3, 7. https://doi.org/10.33735/phimisci.2022.9320

by changing one's representations of time or value, where such changes would increase or lower, for example, one's levels of impulsivity and modulate responses to threat—e.g., if the available time is plenty, then one may cope or avoid threats, rather than freeze or panic (Seo et al., 2019)—or would enhance or lower aversive affective reactions to anticipated negative outcomes for self or others (Crockett et al., 2010). So, if either of the two RL hypotheses I have sketched is correct, then psychedelic therapies may not mainly "induce different forms of self-modelling" (Letheby, 2021a, p. 10), but would mainly change the ways humans—and non-human animals, too—represent time, value and respond to expected punishment.

In summary, one might worry that compared to a RL approach PP lacks genuinely explanatory unifying power and provides us with only a gappy bridge between computational function, neural mechanism and phenomenology. RL modelling foregrounds temporal experience, threat response and valuation as general mental functions that serotonin would modulate and that are impaired in anxiety disorders, depression and addiction. Interpreting the effects of serotonergic psychedelics through the lenses of RL highlights that psychedelic therapy would be effective mainly when it changes patients' maladaptive affective responses to anticipated punishment, temporal representations and value.

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. *Frontiers in Psychiatry*, *4*(47). https://doi.org/10.3389/fpsyt.2013.00047

Boureau, Y. L., & Dayan, P. (2011). Opponency revisited: Competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, *36*, 74–97. https://doi.org/10.1038/npp.2010.151

Carhart-Harris, R. L., & Friston, K. J. (2019). REBUS and the anarchic brain: Toward a unified model of the brain action of psychedelics. *Pharmacological Reviews*, *71*(3), 316–344. https://doi.org/10.1124/pr.118.017160

Clark, A. (2013). The many faces of precision. Replies to commentaries on "Whatever next? Neural prediction, situated agents, and the future of cognitive science." *Frontiers in Psychology*, *4*, 270. https://doi.org/10.3389/fpsyg.2013.00270

Colombo, M. (2014). Deep and beautiful. The reward prediction error hypothesis of dopamine. *Studies in History and Philosophy of Science, part C. Studies in History and Philosophy of Biological and Biomedical Sciences*, *45*, 57–67. https://doi.org/10.1016/j.shpsc.2013.10.006

Colombo, M. (2016). Social motivation in computational neuroscience: Or, if brains are prediction machines, then the Humean theory of motivation is false. In *The Routledge Handbook of Philosophy of the Social Mind* (pp. 336–356). Routledge.

Colombo, M., & Fabry, R. E. (2021). Underlying delusion: Predictive processing, looping effects, and the personal/subpersonal distinction. *Philosophical Psychology*, 1–27. https://doi.org/10.1080/09515089.2021.1914828

Colombo, M., & Heinz, A. (2019). Explanatory integration, computational phenotypes, and dimensional psychiatry: The case of alcohol use disorder. *Theory & Psychology*, *29*(5), 697–718. https://doi.org/10.1177/0959354319867392

Colombo, M., & Wright, C. (2017). Explanatory pluralism: An unrewarding prediction error for free energy theorists. *Brain and Cognition*, *112*, 3–12. https://doi.org/10.1016/j.bandc.2016.02.003

Corlett, P. R., Frith, C. D., & Fletcher, P. C. (2009). From drugs to deprivation: A Bayesian framework for understanding models of psychosis. *Psychopharmacology*, *206*(4), 515–530. https://doi.org/10.1007/s00213-009-1561-0

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*(40), 17433–17438. https://doi.org/10.1073/pnas.1009396107

Colombo, M. (2022). Serotonin, Predictive Processing, and Psychedelics. *Philosophy and the Mind Sciences*, 3, 7. https://doi.org/10.33735/phimisci.2022.9320

Crockett, M. J., Clark, L., & Robbins, T. W. (2009). Reconciling the role of serotonin in behavioral inhibition and aversion: Acute tryptophan depletion abolishes punishment-induced inhibition in humans. *Journal of Neuroscience*, *29*(38), 11993–11999. https://doi.org/10.1523/JNEUROSCI.2513-09.2009

Dayan, P. (2012). Twenty-five lessons from computational neuromodulation. *Neuron*, *76*(1), 240–256. https://doi.org/10.1016/j.neuron.2012.09.027

Dayan, P., & Huys, Q. J. (2009). Serotonin in affective control. *Annual Review of Neuroscience*, *32*, 95–126. https://doi.org/10.1146/annurev.neuro.051508.135607

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*(4-6), 495–506. https://doi.org/10.1016/S0893-6080(02)00044-8

Doya, K., Miyazaki, K. W., & Miyazaki, K. (2021). Serotonergic modulation of cognitive computations. *Current Opinion in Behavioral Sciences*, *38*, 116–123. https://doi.org/10.1016/j.cobeha.2021.02.003

FitzGerald, T. H., Dolan, R. J., & Friston, K. (2015). Dopamine, reward learning, and active inference. *Frontiers in Computational Neuroscience*, *9*(136). https://doi.org/10.3389/fncom.2015.00136

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: Dopamine and decision making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(20130481). https://doi.org/10.1098/rstb.2013.0481

Hoyer, D., Hannon, J. P., & Martin, G. R. (2002). Molecular, pharmacological and functional diversity of 5-HT receptors. *Pharmacology Biochemistry and Behavior*, *71*(4), 533–554. https://doi.org/10.1016/S0091-3057(01)00746-8

Jacobs, B. L., & Fornal, C. A. (1999). Activity of serotonergic neurons in behaving animals. *Neuropsychopharmacology*, *21*(1), 9–15. https://doi.org/10.1016/S0893-133X(99)00012-3

Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, *20*(9), 1293–1299. https://doi.org/10.1038/nn.4615

Letheby, C. (2021a). Naturalistic entheogenics: Précis of philosophy of psychedelics. *PsyArXiv*. https://doi.org/10.31234/osf.io/ztewb

Letheby, C. (2021b). *Philosophy of Psychedelics*. Oxford University Press.

Letheby, C., & Gerrans, P. (2017). Self unbound: Ego dissolution in psychedelic experience. *Neuroscience of Consciousness*, *1*. https://doi.org/10.1093/nc/nix016

Litwin, P., & Miłkowski, M. (2020). Unification by fiat: Arrested development of predictive processing. *Cognitive Science*, *44*(7), 12867. https://doi.org/10.1111/cogs.12867

Liu, Z., Lin, R., & Luo, M. (2020). Reward contributions to serotonergic functions. *Annual Review of Neuroscience*, *43*, 141–162. https://doi.org/10.1146/annurev-neuro-093019-112252

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. https://doi.org/10.1038/nn.2723

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., & Friston, K. J. (2013). Free energy, precision and learning: The role of cholinergic neuromodulation. *Journal of Neuroscience*, *33*(19), 8227–8236. https://doi.org/10.1523/JNEUROSCI.4255-12.2013

Seo, C., Guru, A., Jin, M., Ito, B., Sleezer, B. J., Ho, Y. Y., & Warden, M. R. (2019). Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode. *Science*, *363*(6426), 538–542. https://doi.org/10.1126/science.aau8722

Yon, D., & Frith, C. D. (2021). Precision and the Bayesian brain. *Current Biology*, *31*(17), 1026–1032. https://doi.org/10.1016/j.cub.2021.07.044

Colombo, M. (2022). Serotonin, Predictive Processing, and Psychedelics. *Philosophy and the Mind Sciences*, 3, 7. https://doi.org/10.33735/phimisci.2022.9320