Psychedelics and consciousness

An interview with Robin Carhart-Harris

by Martin Fortier and Raphaël Millière

Citation: Carhart-Harris, R., Fortier, M., & Millière, R. (2017). Psychedelics and consciousness. An interview with Robin Carhart-Harris. *ALIUS Bulletin*, *1*, 1-16.

Robin Carhart-Harris
r.carhart-harris@imperial.ac.uk
Centre for Neuropsychopharmacology
Imperial College, London, UK

Martin Fortier
martin.fortier@ens.fr
Institut Jean Nicod
ENS/EHESS, Paris, France

Raphaël Millière
raphael.milliere@philosophy.ox.ac.uk
Faculty of Philosophy
University of Oxford, UK

What got you interested in the topic of psychedelics?

It was an interest in the mind. At the time (about 12 years ago), I felt that my curiosity about the mind could be best satisfied by a study of psychoanalysis, so I was studying psychoanalysis for a master's degree. Someway in, I began to realize that the main methods of psychoanalysis to access the 'unconscious mind' are unreliable. The existence of the unconscious mind is a cornerstone of psychoanalytic theory - yet its existence is largely denied or at least neglected by mainstream psychology and psychiatry. Dreaming is perhaps psychoanalysis' most valued window in on 'the unconscious' but dreaming happens during sleep and it is easy to forget or confabulate our dream material. Free association also seemed vague and unreliable, so I felt compelled to ask whether anything else had been considered as a way of altering the structure of the mind in such a way that access to 'the unconscious' could be facilitated. At the time, I was somewhat ambivalent about psychoanalysis. It felt like its main tenets had substance but there was also a lot of philosophical noise getting in the way of what I saw as a much more fundamental problem, namely testing and potentially demonstrating that the basic constructs of psychoanalysis, such as *repression*, *the ego*, and *the unconscious mind* 'exist' and have substance in both a psychological and biological sense. I felt psychoanalysis needed to get its act together and better demonstrate these things otherwise it could make no real scientific progress and would instead remain insular and somewhat disconnected from mainstream science. I entertained the idea that a drug might exist that could alter the topography of the mind to aid access to the unconscious and then I discovered Stanislav Grof's book, *Realms of the Human Unconscious:*

Observations from LSD Research, and I thought - here it is! The most logical thing from there seemed to be to take this to brain imaging and measure what's going on in the brain as the unconscious mind is accessed under LSD or a similar drug - and that's basically the idea I took to David Nutt, then Professor of Psychopharmacology at the University of Bristol, as a PhD proposal. Although I've never managed to work on that specific project, I'm a lot closer than I was 12 years ago.

In addition to neuroscience, you studied psychoanalysis. Do you think some of the hypotheses of psychoanalysis on the human mind are relevant for scientific research?

Yes.

It seems fair to say that you are one of the proponents of neuropsychoanalysis (e.g., Carhart-Harris & Friston, 2010).

In a sense, yes.

Some authors (e.g., Ramus, 2013) have worried that the neuropsychoanalytic research program might be a nonstarter, because it amounts to projecting neuroscientific discoveries onto psychoanalytic concepts (but not the other way around).

The relationship should be circular to be healthiest. Arguably, psychoanalysis hasn't made any discoveries for quite some time though. Perhaps what psychoanalysis had to tell us, has already been told, and mostly just by a very select number of people (e.g. Freud, Jung and Klein).

The worry, then, is that in neuropsychoanalysis, neuroscience does inform psychoanalysis, but psychoanalysis hardly ever informs neuroscience.

It may inform or even replace cognitive explanations however. I suppose the point is, psychoanalysis isn't discovering in the same way as neuroscience is, because its 'discoveries' were made many years ago, and through observation but not in a controlled and systematic way (unlike neuroscience).

Sceptics might nonetheless argue that the two domains are not mutually enriching. Do you think this is true?

Not necessarily. I think psychoanalysis can provide enriching explanations for things that cognitive psychology and neuroscience struggle with.

In the same line of skepticism, do you agree that only neuroscience (owing to its scientific research methods) brings new findings on the table?

Yes, these days only neuroscience brings new findings to the table but because I believe psychoanalysis can offer a richer (arguably the richest) account of the human mind and condition, I think it has useful things to say about certain neuroscientific discoveries. I accept it doesn't bring anything new but what it brings that is 'old' is still deeply resonant and informative.

Let's take the example of the 'discovery' of the default-mode network (DMN) for example, around the turn of the millennia. Cognitive neuroscientists have struggled with this system as they can't easily identify what its main function is. It seems to be involved in either too much or nothing specific at all. Some have even challenged whether the approach of studying 'resting-state' brain activity (the approach that helped develop the notion of a 'default-mode') is really meaningful at all or even consistent with what cognitive neuroscience is 'meant to be doing' (Morcom and Fletcher, 2007). Terms like 'stimulus independent thought' been used in reference to the network and its functioning, which are telling more about the priorities of cognitive psychology and its behaviorist counterpart than about the phenomena in question. I think it's more interesting to speculate about and endeavor to sample the nature and content of spontaneous cognition, and tools such as 'experience sampling' (Csikszentmihalyi and Larson, 1987), psychedelics, and neuroimaging may prove to be a powerful combination in this respect. True, the DMN may serve some essential physiological functions (Leech et al., 2014), and granted, rudimentary prototypes of the adult human DMN may be observable in lower species (Mantini et al., 2011) and infants (Doria et al., 2010) but laying the emphasis here seems more about evading the problem than tackling it – and moreover, these things needn't be inconsistent with psychoanalytic conceptions of mind-function – such as the idea that the ego offers a reserve of energy that can be 'invested' in 'objects' of interest (Carhart-Harris et al., 2008).

> "If you want to look for 'the unconscious' in the brain, you probably needn't look far from the DMN and particularly its subcortical/paralimbic components, such as the extended hippocampus."

In 2008 (Carhart-Harris et al., 2008) and later in 2010 (Carhart-Harris & Friston, 2010) and more recently (Carhart-Harris et al., 2014), I proposed that the default-mode network may relate to 'the ego' system, as described by Sigmund Freud. That hypothesis seems to have been quite popular (e.g. the 2010 paper with Karl Friston has been cited over 250 times) and I still feel relatively comfortable about having proposed it. All I would want to do now to qualify it, would be to say that in Freudian theory, in health, 'the unconscious' is largely continuous with the

ego and the ego can contain much that is of 'the unconscious', just as the unconscious contains much of the ego, i.e. they are not so clearly differentiable from each other and exist instead, in relative harmony. It is in ill health that the systems are in conflict. The point I'd like to make here, is that if you want to look for 'the unconscious' in the brain, you probably needn't look far from the DMN and particularly its subcortical/paralimbic component/s, such as the extended hippocampus. If you'd rather stick to cognitive or non-psychoanalytic terms when talking about these systems, e.g. proposing that the DMN is related to 'the narrative self' – then that's fine, I wouldn't want to argue with that – but at least know that psychoanalysis has talked about these systems in depth for over a century (Freud, 1927; Freud, Strachey, & Freud, 2001) and may offer a level of explanation that you might find surprisingly useful and informative. Also, if you're going to judge psychoanalysis and particularly Freudian theory, judge it directly by going to his original work and not to interpretations of it.

How do you think psychedelics research can contribute to the search for the neural correlates of consciousness?

Psychedelics alter 'consciousness' in a marked and novel way. Most studies of consciousness address the problem of consciousness via looking at states of reduced consciousness, e.g. anesthesia, sleep, brain injury and/or illness. Psychedelics do not really reduce the level of wakefulness or alertness but they do fundamentally alter the quality of consciousness. Psychedelics can also help motivate a more nuanced version of the question, 'what are the neural correlates of consciousness?' Because if you look at that question critically, it's pretty vague, i.e. the definition of consciousness is pretty vague. The dominant one is 'consciousness is that which is lost when we fall into dreamless sleep and returns when we wake' but I would challenge that and say instead that what is lost and returns either side of dreamless sleep is actually more or less the whole of the mind, and the whole of the mind is not just consciousness.

What is your general view on the topic of consciousness?

It needs to better define its terms. If the field of 'consciousness research' clarified that the reason we're really interested in consciousness is because we're wonderstruck about the complex nature of *human* consciousness and want to better understand its basis in the brain, then that would be a good start, because then we could focus on humans and how they differ from other animals and perhaps realize that the existence of 'the ego' is quite a fundamental difference. After that, we might realize that rather than aimlessly studying 'consciousness' and approaching it from the same old vantage of reduced states of consciousness, we might better focus on self-consciousness and its neural correlates. A stronger focus on self-consciousness

and the nature of the self as a system (i.e. 'the ego' in psychoanalytic terminology) would help along the initiative of trying to better understand the brain basis of human consciousness – and why human consciousness is special. I think the endeavor to better understand the brain basis of 'consciousness' is too vague a question/problem, and unless I'm missing something (which I may be), I don't think it's going to come to much. It may sound anthropocentric but I think we need to better admit/acknowledge that the real motivator driving enthusiasm for 'consciousness research' is the promise of understanding why and how the human mind is different to that of other animals – and why/how we can reflect on our own thought and behavior. I'm actually open to the idea that it is something relatively recently learned or acquired though, as was argued by Julian Jaynes (1990). Jaynes' stuff about the breakdown of hemispheric separation, I'm less convinced about however, but the idea that (reflective) consciousness is a relatively recent acquisition of the human mind, is a fascinating one – and has some intuitive appeal.

Do you endorse any of the following approaches to consciousness: a topological approach (claiming that there is an area of consciousness), an electrophysiological approach (claiming that consciousness consists of a specific rhythm of neural activity), or a reticular approach (claiming that consciousness consists of a specific connectivity between and within networks)?

If any of those, it would be the last one, but like I said, I think you need to start with a clearer definition of what the problem is, i.e. 'what is consciousness'? Are we talking about the reflective capacity of healthy adult humans? If yes, then I vote for focusing on the DMN and how it interacts with the rest of the brain. DMN-parahippocampal interactions are of particular interest to me in this regard. I wouldn't want to say that the DMN as an object in isolation is consciousness or indeed 'the ego', but I would advise that we study its behavior, and how it develops, how its constituent parts combine to form a gestalt – and how this gestalt interacts with others in the brain, and how the interaction of these gestalts relates to subjective phenomena that we feel and can (mostly) be conscious of, and report.

How do you interpret the seeming discrepancy between Vollenweider et al.'s study (1997) and your own study (Carhart-Harris et al., 2012) regarding the effect of psilocybin on frontal cortex activity?

This is covered in our more recent *PNAS* paper on LSD (Carhart-Harris et al., 2016).

In your seminal paper you pinpoint the difference in recording techniques (PET vs. fMRI) as possibly explaining why one observes such a discrepancy between your findings and Vollenweider's: "this discrepancy relates to the fact that the radiotracer used to measure glucose metabolism (18F-fluorodeoxyglucose) has a long half-life (110

min). Thus, the effects of psilocybin, as measured by PET, are over much greater timescales than indexed by our fMRI measures" (Carhart-Harris et al. 2012, p. 2141). This hypothesis is corroborated by a recent fMRI study of ayahuasca which also detects a decrease in frontal activity (Palhano-Fontes et al., 2015).

Not really, as that mPFC decrease was task-related I think. A team working at Kings College London replicated our arterial spin labeled (ASL) intravenous psilocybin study and found decreased blood flow however. As I say in the 2016 *PNAS* paper, I think the discrepancy is due to the methods. Unless you are *au fait* with the methods, it's easy to think brain imaging literally reads-out 'brain activity', but that isn't at all the case. The process of recording to analysis to brain images, involves a number of assumptions, e.g. people often assume that the BOLD signal of fMRI is measuring brain activity and similarly that the ASL signal represents brain activity – but that's not always the case. For example, a direct vascular action of a drug can interfere with the BOLD and ASL signal and give you a read-out that you misinterpret as being a change in 'brain activity'. I think in the case of the cerebral blood flow (CBF) reductions that have been seen with intravenous psilocybin, it may represent an initial vasoconstriction in the brain but it could also represent an initial change (reduction) in brain activity (when I say 'brain activity', I really mean *neuronal activity*). The overlap between location of the CBF reductions and changes in oscillatory power seen with intravenous psilocybin and MEG was quite remarkable however (Muthukumaraswamy et al., 2013) – and importantly, the latter is a much more reliable and direct measure of brain activity. Basically, if you want this matter to be comprehensively resolved, you need to do a bit more work, perhaps in animals, looking at vascular action of intravenous psilocybin. But I'm not sure it's a problem that is that important to worry about however, as the field has moved on, as have our techniques, and we're now using better measures that more directly sample neuronal activity, and these are yielding reliable results across drugs and study teams. I would basically advise someone worried about this matter, not to worry, as consistent principles about the acute action of psychedelics on the brain are emerging and will be shown to be quite reliable. Our 2016 *PNAS* paper on LSD in the most comprehensive in the sense. In brief, look to the present and future for the answer/s.

> "Increased PFC metabolism is not a good explanation for how psychedelics work to produce their characteristic effects."

In addition to the recording technique, do you think that the mode of administration could also explain the discrepancy: i.e., oral administration in Vollenweider's study vs. intravenous administration in your own?

Possibly. I feel pretty confident that increased PFC metabolism is not a good explanation for how psychedelics work in the human brain to produce their characteristic psychological effects however.

On a related note, what are the potential limitations of fMRI imaging to study psychedelics? More generally, what are the benefits and downsides of each available monitoring techniques to study the activity of the brain on psychedelics?

fMRI doesn't measure brain activity directly. Psychedelics may well have a direct vascular action and this could confound interpretations of some fMRI findings with psychedelics, such as CBF measures. Fludeoxyglucose PET has very poor to no temporal resolution and so provides little/no information about brain dynamics. Simultaneous EEG-fMRI is an important way forward and we're embracing that. Other techniques will likely also emerge in time. For example, the temporal resolution of fMRI is improving – something we might also try to make the most of. Dynamic EEG-fMRI measures twinned with experience sampling may be a powerful way forward. Similarly, decoding methods could prove useful.

Psychedelics researchers have emphasized the importance of set and setting, that is the participant's state of mind and environment when the drug is administered. How do you think set and setting could be modeled and controlled with more precision within experimental studies, given the influence they have on the participant's experience?

It's a challenge because you want to test 'set and setting' as variables but you also want to maintain safety and certain ethical standards. Music offers a good means of modulating 'setting' and 'set' requires that we sample 'where people are at' psychologically prior to the trip itself. This hasn't really been done properly yet – at least by us. You could also try and manipulate expectations to manipulate set – and I suppose we've done that to an extent when looking at suggestibility – but far more work could and should be done here.

What are the difficulties of gathering reliable data about the subjective effects of psychedelics, and how can they be overcome?

Reports are given in retrospect and so can be unreliable and sensitive to biases. Subjective reports are difficult to obtain in real time however, as collecting them will affect the experience and also language skills may be impaired under a potent drug. One potential solution is 'experience sampling'. I'm quite keen to incorporate this into our work. I'm also aware of more sophisticated interviewing techniques

that are being developed - which may be particularly useful. Video and audio recording of sessions and interviews would also be useful.

In order to chart the neurophenomenology of psychedelic experience, the questionnaires you use resort to concepts such as "looking strange" or "having a supernatural quality" and correlations are subsequently made between subjective reports (the feeling that an experience was strange or supernatural) and neurophysiological data (a certain pattern of neural activity). Now, concepts such as that of strangeness or supernaturalness are notoriously ambiguous and likely to be interpreted in various ways by subjects. Thus, when correlations are made between subjective reports and neural patterns, what is being correlated may be extremely different in one case and in another.

To elaborate a little on the example of supernaturalness, it has been shown that this property is variably ascribed (depending on one's personality, one's culture and one's level of expertise) to experiences characterized as: (i) being highly fluent, (ii) being highly disfluent, (iii) being sensorially vivid, (iv) being numinous, (v) being non-dual, (vi) featuring extraordinary beings, (vii) involving a loss of the sense of agency, etc. (see for example: Shanon, 2002; Taves, 2009; Laughlin, 2011; Luhrmann, 2012; Halloy, 2015). This diversity of meanings lying behind the concept of supernaturalness is highly problematic.

Do you think when they are dealing with such polysemous words, neuroscientists should finesse the concepts they use in their questionnaires by consulting anthropologists and phenomenologists who have extensively studied the underlying polysemy of these terms?

I do think we need to better define our terms, yes. Words are certainly vulnerable to interpretation and biases, but language is a difficult prison-house to escape from, however you might try. Words that are especially vulnerable to different interpretations can be problematic and if you want them to be interpreted in a particular way, then you could provide participants with some kind of briefing about what is meant by the term/s in question but then that could bias/prime people to see things 'how you want them to see things', thus causing a confirmation bias.

"Words are certainly vulnerable to interpretation and biases, but language is a difficult prison-house to escape from."

You have investigated a phenomenon known as drug-induced ego dissolution (DIED), which is usually described as a breakdown of one's 'sense of self' and a feeling of unity with one's environment (Lebedev et al., 2015). You have suggested that DIED might be explained as a breakdown of the so-called narrative self, that is the network of beliefs, thoughts and autobiographical memories associated with being the particular person one self-identifies with. This is consistent with subjects reporting feeling as if they were not a person anymore when undergoing DIED during psychedelics use. However, there is room for debate about whether DIED is merely a disruption of the narrative self, and not also of lower-level self-specifying processes, related to the so-called « minimal » or « bodily » self. There is a long tradition in philosophy of discussing the idea of the minimal self (also called « sense of mineness ») as the pre-reflective, nonconceptual feature of consciousness in virtue of which my experiences feel mine (Zahavi, 2014). This notion has also come under investigation in cognitive science and psychiatry. According to neurocognitive accounts of the minimal self, especially in a predictive coding framework, it is crucially linked to multisensory integration (particularly of visuotactile and vestibular input), interoception and homeostatic regulation (Christoff et al., 2011; Limanowski & Blankenburg, 2013; Apps & Tsakiris, 2014). From a phenomenological point of view, it might be reducible to body ownership, self-location, and the experienced direction of the first-person perspective (Blanke & Metzinger, 2009). The minimal self does not depend on high level cognitive processes such as self-related beliefs and first-person thoughts, but rather on low-level bodily/perceptual processes. In this framework, the minimal self is a necessary condition for reflective self-consciousness (self-related thoughts), which in turns enables self-related beliefs (narrative self). If this is right, one could expect the narrative self to break down if the minimal self is disrupted, and this is indeed what appears to be the case to some extent in schizophrenia.

The idea that DIED is primarily a disruption of the minimal self seems consistent with the fact that classical psychedelics induce hallucinations and not delusions. In a predictive coding framework, given that drug-induced hallucinations presumably stem from impaired bottom-up processing coupled with relatively preserved top-down processing, it seems reasonable to expect excessive prediction error signaling to yield aberrant predictions and a disruption of multisensory integration, ultimately leading to the breakdown of the minimal self. In other words, DIED would primarily be about a perceptual/bodily anomalous processing that leads to the feeling to being selfless, rather than a cognitive anomaly leading to delusions impairing self-narratives.

Do you think that such an account of DIED has any plausibility? How might further studies bear on this debate?

The ego, in the Freudian sense, incorporates the narrative and bodily self. I understand the differences between these different aspects of 'self' but I'm not sure how important it is that we chop them up in order to understand DIED. I tend to think when people rate DIED they are recognizing that it is 'the ego' in the Freudian sense that is compromised, i.e. the self as a *system*. Generally speaking, the way people understand and use the term 'the ego' in everyday parlance, is more or less consistent with Freud's account of it. Our ego-dissolution inventory puts emphasis on reduced 'self-importance' and a sense of connectedness or oneness (e.g. to self, others and nature), which naturally accompanies DIED. I suppose what's happening is that boundaries necessary for the existence of the ego, breakdown and the sense of connectedness is the inevitable result. People can fight DIED but then they won't feel the connectedness.

"I tend to think when people rate DIED they are recognizing that it is 'the ego' in the Freudian sense that is compromised."

The DMN is usually associated with mind-wandering, self-reflection and introspection while the Task Positive Network (TPN) is associated with the orientation of the mind towards the external world. Elaborating on this distinction, you notice that these two networks are anti-correlated in order to guarantee the functionally important distinction between the internal world and the external one. Now, with your colleagues, you have shown that the DMN/TPN anti-correlation is significantly decreased after psilocybin intake. You have proposed that the diminution of the orthogonality between the DMN and the TPN was a very plausible explanation as to why psychedelic experiences are characterized by a "collapse of dualities" and by the "disturbance in one's sense of self, and particularly one's sense of existing apart from one's environment" (Carhart-Harris et al., 2014, p. 16). Your proposal seems very intuitive indeed: if the DMN and the TPN stop being strongly anti-correlated this should result in a loss of neat distinction between the internal and the external.

However, there is good evidence suggesting that the loss of the internal/external dichotomy is to be explained by something different from the disturbance of the DMN/TPN anti-correlation. Indeed, the "collapse of dualities" and the "disturbance in one's sense of self" seem to characterize the phenomenology of both psilocybin-induced experiences and ayahuasca-induced ones. Now, given that ayahuasca experiences also involve some confusion between the internal and the external, we would expect to find a clear decrease of the DMN/TPN anti-correlation after ayahuasca

intake. Remarkably enough, this doesn't seem to be the case (Palhano-Fontes et al., 2015). How do you interpret these findings?

The specifics of the data processing and statistical thresholds can explain Palhano-Fontes's alleged negative result. If you look closer at the method they used however, I think they did find this effect (i.e. reduced DMN-TPN orthogonality) – and I'm very confident others will in the future.

Do you think that it can still be maintained that the loss of a clear dichotomy between the internal and the external is to be explained by the decrease of the DMN/TPN anti-correlation?

Yes. Although we don't need to be too specific about the TPN, as there are a few "TPNs". "Task positive network" is a bit of a vague term for a network to be honest.

On a more methodological note, there has been some intense debate as to whether regressing out the global signal is a sound way of measuring anti-correlation between networks (Fox et al., 2009; Murphy et al., 2009). If regressing out the global signal turns out not to be valid, then the DMN/TPN anti-correlation could simply be an artefact. What is your take on this methodological issue?

We did our most recent analysis (*PNAS* 2016) with and without global signal regression and found the same result. This is discussed in our paper. You just have to be careful about the terms you use and, for example, go with "orthogonality" instead of "anti-correlation". I think this is discussed in Leor Roseman's *Human Brain Mapping* paper (2016).

In your 2014 Frontiers paper you sketch a general model of Altered States of Consciousness (ASCs) which aims at theorizing not only psychedelic states but more broadly any kind of altered or anomalous state (such as psychosis, coma, dreaming, etc.). Your proposal is that entropy is a spectrum which can take different values: for example, early psychosis is characterized by high entropy whereas coma or sedation are characterized by low entropy. Your model is also based on another dimension: that of criticality (i.e., the ability of the brain to reach certain critical thresholds beyond which new complex properties emerge, notably through cascade-like processes) (e.g., Beggs & Plenz, 2003; Chialvo, Balenzuela, & Fraiman, 2008). Interestingly, in your model, you defend the view that entropy and criticality are closely correlated: the highest the entropy, the highest the criticality; conversely, the lowest the entropy, the lowest the criticality.

Did I say that? Criticality is just one thing, i.e. it's a critical point, so you can't really have low or high criticality – but you can be above or below it. Being super-critical,

in the sense of being *above* a critical point, would be most consistent with a more entropic (random) state.

Don't you think that there are cases in which entropy can vary independently of criticality? For example, it has been proposed that certain altered states can involve plenty of prediction errors while the weight (or the accuracy) ascribed to these predictions errors remains abnormally low; conversely, predictions errors can be low while the accuracy ascribed to these limited predictions errors is very high. As Fletcher and Frith put it, "a relatively small prediction error might be given undue weight (if the uncertainty is underestimated), leading to a false inference. Alternatively, excessive noise might dilute the effects of even a large prediction-error signal, leading to a reluctance to accept an inference as adequately explaining the input." (Fletcher & Frith 2009, p. 55) Let's take the example of a state characterized by high prediction errors and low accuracy. In such a case, priors couldn't be revised and updated as they should and they would end up being abnormally steady (e.g., Fletcher & Frith, 2009; Adams, Brown, & Friston, 2015). The brain would thus combine both high entropy (there are plenty of prediction errors) and sub-criticality (high-level mental states are not malleable). This seems to contradict the idea that high entropy and high criticality always go hand in hand.

They don't, as you can't have high criticality, but you have "super-criticality" – and generally speaking, though perhaps not absolutely, super-criticality would be consistent with high entropy.

In your work, you have occasionally pondered upon what your findings suggest as to what the broad structure of the brain is. You have notably advanced that two models of the brain can shed very interesting light on psychedelics: the "free-energy principle" put forth by Karl Friston (2010; 2006) and the "reducing valve" model proposed by Aldous Huxley (1954). In your seminal study you seem to suggest that these two models are in fact almost identical (Carhart-Harris et al. 2012, p. 2142).

Huxley's model was explicitly inspired by Charlie D. Broad (1953, chap. 1), and Broad himself borrowed the "reducing valve" metaphor from Henri Bergson (1994 [1896]). Now, if we look at it closely enough, the model endorsed by Bergson, Broad and Huxley is arguably quite different from that defended by Friston. Bergson's key idea is that the default state of the brain is that of being overwhelmed by sensory data coming from the world; the brain's function is thus to diminish the amount of sensory data reaching consciousness. This view straightforwardly contradicts Kant's transcendental idealism: by and large, Kant says that the data coming from the world are poor and that the brain later enrich them; on the contrary, Huxley, Broad and Bergson claim that the data coming from the world are too rich to be processed by a normal brain and that the brain is precisely there to filter the massive surge of data coming from the world.

In this regard, Friston's model seems to fundamentally differ from that advanced by Bergson and his followers. The free-energy and predictive coding framework has it that prediction comes first and prediction errors (sensory data coming from the world) come next to rectify and update top-down predictions. While in Bergson's model the brain is simply a filter which cannot by itself generate any conscious representation (the richness of representations is to be found in the world), in Friston's model, the brain is notoriously able to generate conscious representation and it is even able to do so before getting any sensory feedback. In other words, the free-energy and predictive view contends that the role of the world is corrective and that the role of the brain is constitutive whereas Bergson, Broad and Huxley hold exactly the opposite view: for them, the world is constitutive and the brain is simply corrective.

Do you agree with the distinctions drawn here between Bergson's model and Friston's model? Five years after your seminal study, do you tend to think that the "reducing valve" model provides the best account of psychedelics or that the "predictive" model does a better job?

Friston is a leading contemporary neuroscientist whose free-energy principle is an elegant, empirically-informed model of how the brain works. Huxley was a brilliant author and philosopher but not a neuroscientist. The reducing valve idea is quite nice as a metaphor but I think people take it too literally and sometimes even want to use it in a sort of pseudo-scientific way, to suggest that there is a filter that stops us seeing what's really "*out-there*" in a matrix-esque kind of way.

If the metaphor is useful however, it's useful because it proposes that "the brain, in main, is eliminative rather than productive". That idea is consistent with the free-energy principle because top-down inferences work to explain bottom-up sensory information – so there's some functional suppression going on but suppression in the sense of predictive processing. Basically, I wouldn't worry too much about Bergson, Broad and Huxley or indeed Kant when it comes to a contemporary account of how the brain works. It's best to see where the field is now, and Friston's free-energy model is one of the best the field has.

In your opinion, what are the next steps for psychedelics research?

There's so much but better predicting response to the psychedelics is a good example of one potentially fruitful area.

References

Adams, R., Brown, H., & Friston, K. (2015). Bayesian inference, predictive coding and delusions. *Avant: Trends in Interdisciplinary Studies*, *5*(3).

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews*, *41*, 85–97.

Beggs, J. M., & Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *The Journal of Neuroscience*, *23*(35), 11167–11177.

Bergson, H. (1994 [1896]). *Matter and memory*. New York: Zone Books.

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13.

Broad, C. D. (1953). *Religion, Philosophy and Psychical Research*. London: Routledge / Kegan Paul.

Carhart-Harris, R., Erritzoe, D., Williams, T., Stone, J. M., Reed, L. J., Colasanti, A., … Nutt, D. J. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proceedings of the National Academy of Sciences*, *109*(6), 2138–2143.

Carhart-Harris, R., & Friston, K. (2010). The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain: A Journal of Neurology*, *133*, 1265–1283.

Carhart-Harris, R., Leech, R., Hellyer, P., Shanahan, M., Feilding, A., Tagliazucchi, E., … Nutt, D. (2014). The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs. *Frontiers in Human Neuroscience*, *8*.

Carhart-Harris, R., Mayberg, H., Malizia, A., & Nutt, D. (2008). Mourning and melancholia revisited: correspondences between principles of Freudian metapsychology and empirical findings in neuropsychiatry. *Annals of General Psychiatry*, *7*, 9. 9

Carhart-Harris, R., Muthukumaraswamy, S., Roseman, L., Kaelen, M., Droog, W., Murphy, K., … Nutt, D. (2016). Neural correlates of the LSD experience revealed by multimodal neuroimaging. *Proceedings of the National Academy of Sciences*, *113*(17), 4853–4858.

Chialvo, D., Balenzuela, P., & Fraiman, D. (2008). The brain: What is critical about it? *Conf. Proc. Am. Inst. Phys.*, *1028*, 28–45.

Christoff, K., Cosmelli, D., Legrand, D., & Thompson, E. (2011). Specifying the self for cognitive neuroscience. *Trends in Cognitive Sciences*, *15*(3), 104–112.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *The Journal of Nervous and Mental Disease*, *175*(9), 526–536.

Doria, V., Beckmann, C., Arichi, T., Merchant, N., Groppo, M., Turkheimer, F., … Edwards, D. (2010). Emergence of resting state networks in the preterm human brain.

Proceedings of the National Academy of Sciences of the United States of America, *107*(46), 20015–20020.

Fletcher, P., & Frith, C. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58.

Fox, M. D., Zhang, D., Snyder, A. Z., & Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *Journal of Neurophysiology*, *101*(6), 3270–3283.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology - Paris*, *100*(1-3), 70–87.

Halloy, A. (2015). *Divinités incarnées: L'apprentissage de la possession dans un culte afro-brésilien*. Paris: PETRA.

Huxley, A. (1954). *The Doors of Perception and Heaven and Hell*. London: Harper & Brothers.

Jaynes, J. (1990). *The origin of consciousness in the breakdown of the bicameral mind*. New York: Houghton Mifflin.

Laughlin, C. D. (2011). *Communing with the Gods: Consciousness, culture and the dreaming brain*. Brisbane: Daily Grail.

Lebedev, A. V., Lövdén, M., Rosenthal, G., Feilding, A., Nutt, D. J., & Carhart-Harris, R. L. (2015). Finding the self by losing the self: Neural correlates of ego-dissolution under psilocybin. *Human Brain Mapping*, *36*(8), 3137–3153.

Leech, R., Scott, G., Carhart-Harris, R., Turkheimer, F., Taylor-Robinson, S., & Sharp, D. (2014). Spatial Dependencies between large-scale brain networks. *PLOS ONE*, *9*(6).

Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, *7*.

Luhrmann, T. (2012). *When God talks back: Understanding the American evangelical relationship with God*. New York: Alfred Knopf.

Mantini, D., Gerits, A., Nelissen, K., Durand, J.-B., Joly, O., Simone, L., … Vanduffel, W. (2011). Default Mode of Brain Function in Monkeys. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(36), 12954–12962.

Morcom, A., & Fletcher, P. (2007). Does the brain have a baseline? Why we should be resisting a rest. *NeuroImage*, *37*(4), 1073–1082.

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., & Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *NeuroImage*, *44*(3), 893–905.

Muthukumaraswamy, S., Carhart-Harris, R., Moran, R., Brookes, M., Williams, T., Errtizoe, D., … Nutt, D. (2013). Broadband cortical desynchronization underlies the human psychedelic state. *Journal of Neuroscience*, *33*(38), 15171–15183.

Palhano-Fontes, F., Andrade, K. C., Tofoli, L. F., Santos, A. C., Crippa, J. A. S., Hallak, J. E. C., … Araujo, D. B. de. (2015). The psychedelic state induced by ayahuasca modulates the activity and connectivity of the default mode network. *PLOS ONE*, *10*(2), e0118143.

Ramus, F. (2013). What's the point of neuropsychoanalysis? *The British Journal of Psychiatry*, *203*, 70–71.

Roseman, L., Sereno, M., Leech, R., Kaelen, M., Orban, C., McGonigle, J., … Carhart-Harris, R. (2016). LSD alters eyes-closed functional connectivity within the early visual cortex in a retinotopic fashion. *Human Brain Mapping*, *37*(8), 3031–3040.

Shanon, B. (2002). *The antipodes of the mind: Charting the phenomenology of the ayahuasca experience*. New York: Oxford University Press.

Taves, A. (2009). *Religious experience reconsidered: A building-block approach to the study of religion and other special things*. Princeton NJ/Oxford: Princeton University Press.

Vollenweider, F., Leenders, K., Scharfetter, C., Maguire, P., Stadelmann, O., & Angst, J. (1997). Positron Emission Tomography and Fluorodeoxyglucose Studies of Metabolic Hyperfrontality and Psychopathology in the Psilocybin Model of Psychosis. *Neuropsychopharmacology*, *16*(5), 357–372.

Zahavi, D. (2014). *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford University Press.