# Bayesing Qualia: consciousness as inference, not raw datum

Article (Accepted Version)

# Bayesing Qualia: Consciousness as Inference, not Raw Datum

**Andy Clark[1], Karl Friston[2], and Sam Wilkinson[3]**

## Abstract

The meta-problem of consciousness (Chalmers (this issue)) is the problem of explaining the behaviors and verbal reports that we associate with the so-called 'hard problem of consciousness'. These may include reports of puzzlement, of the attractiveness of dualism, of explanatory gaps, and the like. We present and defend a solution to the meta-problem. Our solution takes as its starting point the emerging picture of the brain as a hierarchical inference engine. We show why such a device, operating under familiar forms of adaptive pressure, may come to represent some of its mid-level inferences as especially certain. These mid-level states confidently re-code raw sensory stimulation in ways that (they are able to realize) fall short of fully determining how properties and states of affairs are arranged in the distal world. This drives a wedge between experience and the world. Advanced agents then represent these mid-level inferences as irreducibly special, becoming increasingly puzzled as a result.

[1] University of Sussex, Brighton, UK, and Macquarie University, Sydney, Australia
[2] Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London (UCL), London, United Kingdom
[3] University of Exeter, UK

## 1. Methodological preliminaries

The 'hard problem of consciousness' is the problem (Chalmers (1996)) of explaining how physical events give rise to the varieties of conscious phenomenal experience. The meta-problem of consciousness (Chalmers, 2018) is the problem of explaining why we think there is a hard problem in the first place. It is the problem of explaining why it is that some intelligent agents find themselves deeply puzzled by certain features of their own contact with the world - puzzled enough, in some cases, to announce the existence of a profound 'explanatory gap' between their best imaginable scientific grip upon how physical things work and the nature and origins of their own experience.

Care is needed in setting up the meta-problem. We need to understand the meta-problem in a way that is (broadly speaking) behavioral rather than making essential reference to phenomenal experience itself. In practice, this means the goal is to explain the things we say and do, while bracketing the question of whether or not they reflect phenomenal experience. Specifically, 'meta-problem' apt behaviors would thus include saying things such as 'there is a profound explanatory gap separating my phenomenal experience and good scientific explanation', and expressing puzzlement about 'qualia' – about why red looks the way it looks, or why pains feel the way they do, or feel like anything at all. Chalmers 2018 thus describes the situation as one in which what we seek to explain are "dispositions to make quasi-phenomenal reports, where reports are understood as outputs that even a non-conscious being could make". In just this vein we aim to show why a certain kind of inference machine will be led to

conclude that it is home to some very puzzling states that have many of the hallmarks of 'qualia'. In so doing we hope, moreover, to lay the groundwork for a substantive, but revisionary, account of consciousness itself.

Which brings us to our title 'Bayesing Qualia'. That title pays homage to Dennett (1988) who, both in *Quining Qualia* and in subsequent work (e.g. Dennett (2015)), has argued that qualia involve some kind of illusion. In 'Quining Qualia' the response to that illusion was to follow Quine in eliminating such misleading posits from (at least) the scientific image. But in what follows we aim not to Quine (explain away) qualia but to 'Bayes' them – to reveal them as products of a broadly speaking rational process of inference, of the kind imagined by the Reverend Bayes in his (1763) treatise on how to form and update beliefs on the basis of new evidence. Our story thus aims to occupy the somewhat elusive 'revisionary' space, in between full strength 'illusionism' (see below) and out-and-out realism. If we are right, we do not infer that we have qualitative experiences *because* we see red, feel pain etc. Rather, seeing red and feeling pain (just like seeing dogs, cats, vicars, and even (Letheby and Gerrans (2017)) having a sense of self) are **t**hemselves inferred causes, constructed to accommodate (i.e., best explain) the raw sensory flux – and the hierarchical machinations they induce. But they are inferred causes that are also represented as especially certain. It is this sense of certainty, we argue, that opens up the space for Cartesian puzzlement and the belief in an explanatory gap.

Our account follows Dennett in denying that qualia are just what they seem – raw givens on the basis of which we infer stuff about the world. On our account (like Dennett's) there simply is no such thing as raw experience.

Instead, our brains construct qualia as 'latent variables' – inferred causes in our best 'generative model' (more on that later) of embodied interactions with the world. But thus constructed qualia, we argue, are of a piece (modulo that added certainty, more on which later) with other inferred variables such as dogs, cats, heatwaves, and vicars. This gives our story its slightly more realist tinge. Qualia – just like dogs and cats – are part of the inferred suite of hidden causes (i.e., experiential hypotheses) that best explain and predict the evolving flux of energies across our sensory surfaces.

## 2. Encountering a World

Our starting point is 'predictive processing' (PP) - a simple but powerful approach to perception, action, and learning[4]. PP depicts the biological brain as an evolved organ that continuously tries to predict the next states of its own sensors, using well-understood optimization methods to steadily improve its guesses. Such a process results in the installment of a probabilistic generative model of the distal causes (sometimes called 'hidden causes' or 'latent variables') that might be causing the sensory flux. For example, a system training on lots of sentences in a public language might be led to posit the existence of distinct classes of linguistic entity, such as verbs and nouns, each of which make certain kinds of sentential unfolding much more probable than others. Such a system has, to a first approximation, inferred the existence of *verbs* as a hidden cause of some of the regularities (compressible patterns) found in the sensory stream.

[4] See Friston (2005). For introductions see Clark (2013), Hohwy (2013), Clark (2016a))

When this kind of learning takes place in a multi-level architecture, lower levels discover patterns at shorter scales of space and time, while higher levels use those patterns as the basis for learning about still other patterns, spanning greater scales of space and time (Hohwy (2013))(Murray et al., 2014, Cocchi et al., 2016, Friston et al., 2017). Equipped with a good predictive (generative) model, these systems deliver not just learning but also online perception by the same process of minimizing 'prediction error', where prediction error is simply the difference between current predictions and the sensory evidence at hand. Finally, PP systems that can act upon their worlds can use those actions to bring about patterns of sensory stimulation, thus shaping the sensory stream to fit, test, and update their own predictions (Friston, Rigoli et al (2015)).

It has recently been suggested (Seth (2013) (Clark (2017) Barrett (2017)) that it is the constant inflection of outward-looking predictions by changing bodily information that explain much of the 'embodied feel' of experience. Courtesy of that constant background inflection we encounter a world that is subtly permeated at all times by a sense of the bodily consequences of our own possible or unfolding actions. This delivers a predictive grip on multi-scale structure – in the external world – superimposed upon a second multi-layered predictive grip reporting on the changing physiological state of the body[5].

Putting this all together delivers our starting point. For what we have just described is an organismal form that will use both interoceptive and exteroceptive sensory information to infer important features of its own body and world. Nothing thus far, however, speaks directly to the issues concerning

[5] Dennett (2015) argues for a closely related picture in which 'qualia' are disguised appreciations of our own predictions concerning our reactive dispositions (to approach, avoid, say 'oh that's a cute baby' etc.). See also Clark (2017) and Clark (2016b).

phenomenal experience or the meta-problem of explaining our various Cartesian puzzlements. What we have described is just a robot that can learn about compressible (hence predictable) patterns at multiple scales of space and time, and that can use those patterns to predict and control its own evolving sensory stream. Such a robot has the admirable capacity to recognize and preferentially seek out worldly environments conducive to its own survival and flourishing. But perhaps there need be nothing it is like to be that robot. Nor is that robot yet poised to make what Chalmers called 'quasi-phenomenal reports', or to express (quasi-express) puzzlement concerning its own 'experience', or to intuit (quasi-intuit) the existence of an explanatory gap. More is needed. But what?

## 3. Imaginary Foundations?

Schwarz ((2018) (this issue) suggests that a Bayesian perceiver, in order successfully to conditionalize beliefs upon incoming sensory evidence (where that means transduced energies) might be forced to extend her probability space by adding a kind of new high-certainty dimension – an 'imaginary foundation'. Schwarz develops his story as an ambitious alternative to standard ways of understanding Bayesian inference in hierarchical, multi-level settings. We think (although we will not argue for this today) that this is a mistake. Nonetheless, his picture draws attention to something important; namely, that in perception, we seem to become highly confident of *something*, where that something does *not quite* mandate high-level beliefs about the state of the distal world itself. Thus, to take the case that Schwarz uses to kick off his (2018) treatment, we can't be sure that what we see – when we look out of the

window and see a fountain – is water. It might be vodka instead. Indeed, it might be nothing at all – we might be dreaming or hallucinating the whole thing. In a lucid dream, we could even judge this to be the case. Yet we would remain very certain of *something*, where that something is, intuitively speaking, a bunch of experienced phenomenal features.

Schwarz' imaginary foundations are purpose-built to fill that role. They are purpose-built to be known with great certainty, while not themselves being made true simply by states of the distal world.. Creatures thus equipped would be able, were they sufficiently intelligent, to assert that *despite* holding all the phenomenal facts fixed, how the world really is might vary, even to the point of there being nothing at all bearing the properties so confidently represented as being present. For such beings, Cartesian doubt is possible. Such creatures would also be capable (general intelligence permitting) of important new forms of counterfactual reasoning. The very fact we can entertain hypotheses like "what would I see if water was vodka" tells us an enormous amount about our capacity for counterfactual inference and hypothesis building. Science itself might reasonably be thought to depend upon just these kinds of capacities.

We suggest that 'imaginary foundations', far from being a highly speculative addition to standard accounts of hierarchical Bayesian inference, are in fact a direct consequence of them. They arise when mid-level re-coding[6] of impinging energies are estimated as highly certain, in ways that leave room for the same mid-level encodings to be paired with different higher-level pictures, including

[6] Mid-level re-codings are taken here to be posterior beliefs at intermediate levels of a hierarchical generative model, where beliefs are used in the sense of Bayesian belief updating or belief propagation – more on this later.

ones in which nothing in the world corresponds to the properties and features at all (as we might judge in the lucid dreaming case).

It's not hard to see why evolved creatures might benefit from high mid-level certainty. A creature – whose brain assigns extremely high confidence to signals suggesting imminent tissue damage – is a creature that will act upon those signals without further reflection. The redness of the predator's beak, likewise, should be processed fast and with high enough confidence to recruit immediate evasive action. But as the depth and reach of our generative models increased, we became aware that things are not always as they seem. We became able, amazingly, to deliberately explore multiple possible scenarios consistent with the stuff known with such mid-level certainty. That red beak might be part of a Virtual Reality illusion. We might have been drugged or duped. At that point, we became creatures armed with new ways of understanding their worlds – ready to do science, ready to do philosophy (Cartesian doubt and all), and ready to become increasingly puzzled by our own mid-level certainties. That puzzlement finds its fullest expression in the literature concerning the 'explanatory gap', where we are almost fooled into believing that there's something special about qualia – that they are not simply highly certain mid-level encodings optimized to control adaptive action.

Commenting on an earlier version of this paper, Chalmers (personal communication) writes "at various …points you give reasons for why we should be more confident in qualia or perceptual states than in the external world -- but "more confident" is a long way short of "unshakable certainty"". Here, it is important (as it always is when dealing with these accounts) to clearly distinguish agentive certainty from its probabilistic underpinnings in the whirl

of processing. In order to deal effectively with our worlds on the basis of sensory information, it may have been fruitful for evolution to induce a kind of threshold effect, whereby the agent (the reporting system) feels 100% certain whenever the underlying processing settles into states of extremely high confidence. (For technical reasons it is usually not a good idea for the processing itself to reach 100% certainty as this locks solutions into place in a way that blocks ongoing multi-level processing). So, when mid-level processing reaches some – doubtless context-sensitive – threshold, the agent will feel 100% sure that she is having such and such an experience. This is all that is needed for our story to get a grip. Moreover, since it is only the mid-level states that have reached this threshold, that same agent will be able to reflect somewhat skeptically about her own sensory certainty. She might (to borrow an example suggested by a helpful reviewer) be reading this very paper and hence be led to think the thought that *I should not really have such full certainty about these sensory encodings*. Such an agent recognizes why she cannot seem to doubt her own qualitative experience in the same way she can doubt other things. In recognizing this, she finds herself unable to fully commit to qualia realism (because the mid-level encodings are really no different to other inferred states) while remaining forced to experience the world *as if* those states were known with total certainty.

To bring our picture into full focus, however, we now add a crucial, part of the predictive processing architecture that was omitted from our earlier sketch. That part is the so-called *precision* with which successful intermediate level predictions are currently held. Precision, in these accounts, has been equated with the psychological construct of attention (Feldman and Friston 2010,). To attend to something increases the precision, confidence or certainty invested in

that thing (Parr and Friston 2017), usually at the expense of other sources of evidence. For example, if I am sitting in the dark palpating a mug-shaped object, I will attend to tactile and haptic cues but not visual cues, becoming confident of features of shape and feel. Furthermore, I will know (report) that I am now fairly certain that this is a mug. I will also have (a possibly subpersonal) belief that I have absolutely no confidence in my beliefs about its color. Instead, I know, with great certainty that there is an irresolvable ambiguity in respect of some mysterious property, whose behavior in other real-world situations is captured by the communally handy concept of 'color'. Importantly, both the properties themselves and their degrees of certainty (i.e., precision) are computed by entirely agent-opaque means. The agent can also become aware of other ways the world might be, that are consistent with holding these elusive properties firm. So, she can ask herself what it would look like if it was vodka rather than water in the fountain, or if she was having a dream about a fountain, and realize it would look just the same in all those cases.

At this point, sufficiently intelligent systems may infer the existence of mysterious intervening qualia. Practically speaking, they are warranted to do so. For qualia thus posited prove extremely useful, enabling us better to predict our own and others' future responses. As Dennett (2015) nicely argues, qualia now pass the 'Bayesian test' for presenting genuine, yet somehow strangely elusive, aspects of the world. From the PP perspective, they are just more predictively potent mid-level latent variables in our best generative model of our own embodied exchanges with the world. They are not some kind of raw datum on which to predicate inferences about the state of body and world. Rather, they are themselves among the many products of such inference.

In one way, this is a version of what Frankish (2016) calls 'illusionism'. If the term 'qualia' is constrained to pick out some kind of raw experiential data, then qualia are an illusion, and we only think (infer) that such states exist, But in another sense, this is a way of being a revisionary kind of qualia realist, since colors, sights, and sounds are revealed as generative model posits pretty much on a par with representations of dogs, cats, and vicars. We return to this issue later in our treatment.

## 4. Making It Real

All this unpacks gracefully in the modern setting of hierarchical Bayesian inference. Hierarchical inference is Bayesian belief updating under a hierarchical generative model in which probabilistic beliefs at one level depend upon beliefs at a higher level. All intermediate levels in hierarchical inference now play the role of *empirical* priors; namely, prior beliefs that depend upon the bedrock sensory evidence[7].

An important aspect of these hierarchical models is the scope they provide for a factorization of model or hypotheses space, so as to enable useful conjunctive constraints on lower-level inference . For example, I could have one dimension of model space that entertained two hypotheses: 'it is snowing' and 'it is not snowing'. A second hypotheses space could pertain to the nature of the snow: 'real' versus 'synthetic', or even 'dreamt' (in a lucid dream state). Crucially, it is possible to have precise or definitive beliefs along one factor (e.g., "I am absolutely certain it is snowing") but to be uncommitted over another (e.g.,

[7] Crucially, hierarchical inference necessarily induces beliefs about beliefs. These beliefs include 'qualia' as a potential explanation for lower-level beliefs and, on the present view, qualia *per se*. Equipping a generative model with an extra level means one can entertain (i.e., infer) beliefs about – or explanations for – qualia (e.g., "this is an illusion").

"This could be real or synthetic snow").

Perhaps the most canonical example of hierarchical inference is when the higher level comprises a space of models or hypotheses that establish plausible contexts for inference at the level below. For example, I could entertain two hypotheses that constrain my inference about sensory evidence in some context: "I could be sitting in my front parlor" or "I could be sitting on a film set". If I see white flakes floating down outside my window, my perceptual inference will be profoundly different under the two models (i.e., it is snowing – or someone is using a synthetic snow machine). Crucially, the empirical priors afforded by the second level of my generative model not only constrain my perceptual synthesis but are also informed by higher and lower level beliefs. For example, if I know it is summer (i.e., higher empirical priors) I will assign greater credence to the 'film set' hypothesis over a 'winter snowscape'. Furthermore, if I see that the snowflakes do not melt when settling on warm surfaces (i.e., lower empirical priors), this will reaffirm the 'film set' hypothesis. As noted above, if we subscribe to a deep or hierarchical form of belief updating in our brains, then this lends us a remarkable capacity: namely, I can entertain alternative (counterfactual) models or hypotheses and effectively ask "what would this look like if I was in this situation". In the lucid dream setting, we might reflect that the sensorium would appear *just like this*. At that point, we have opened up the whole space of Cartesian doubt, and sowed the seeds for thinking that experience is deeply special, somehow floating free of states of the real world.

The nature of the processing involved is (from a PP perspective) now clear

enough. It involves the deliberate control of the precisions assigned to various low and high-level beliefs[8]. The advanced perceiver may, for example, forcibly assign high-precision to the 'vodka fountain' belief, so as to become aware that that belief is actually consistent with the current (highly certain) set of mid-level sensory evidence – the evidence that normally supports a 'water-fountain' conclusion. Under such conditions, we explicitly understand that other states of the real world might nonetheless have given rise to the very same sets of incoming sensory stimulation. This confers huge cognitive benefits, plausibly including (as mentioned above) the pursuit of science itself. In less advanced creatures such complex counterfactual probing may not be possible. For them, their own 'mid-level foundations' are never held in focus, while deliberately varying their own higher level beliefs. Such creatures will still conditionalize their top-level beliefs upon simplified, stable, mid-level foundations. But they will not begin to make an appearance/reality distinction or become puzzled by their own qualitative experiences.

What about experiences, whose sensory qualities are different, dilute or less marked? Examples might include some kinds of deliberate imagination, simple thinking, pain, and some emotional states. These are not our focus today, but similar remarks apply across the board. Qualitative states of all these kinds are, we conjecture, underpinned by systemic estimations of high certainty (precision) that present themselves to the agent as being completely certain. And here too, advanced agents are able to reflect that such certainties are in

[8] This control is a subtle but fundamentally important aspect of hierarchical inference: first, it requires hierarchical inference in the sense that precision or certainty has itself to be inferred (c.f., endogenous attention). Second, equipping a generative model with the capacity to infer its own confidence – about mid-level beliefs – eludes phenomenal opacity, in a way that may underwrite the "puzzlement" that attends our sentient prowess.

fact consistent with multiple ways the world and body night really be. In the case of psychogenic pain, for example, a subject may come to believe (perhaps by reading recent accounts of the predictive roots of psychogenic pain, such as Bergh et al (2017),) that the real problem lies not in the gross body but in aberrant patterns of attention and false estimations of precision. Where the qualitative states seem less distinct (as in some exercises of deliberate imagination) this reflects only the levels at which highly certain information is being generated and manipulated – the higher levels having less (and less highly certain) sensory involvement. Our sense that thinking has less of a 'feel' than seeing plausibly reflects only this mundane fact.

**5. Can Bayes-ed Qualia Stand the Strain?**

A natural worry about the story on offer is that it may seem to replace the actual *experience* of qualia with judgments of one form or another – for example, the judgment that I am now seeing a red cup or feeling a sharp pain. Chalmers (2018 p. 9) raises just this kind of worry, noting that on his view "consciousness is real, and explaining our judgments about consciousness does not suffice to solve or dissolve the problem of consciousness". As it stands this is not an argument so much as an assertion of faith. However, the same could be said of our own assertion that our intuitions concerning qualia can be fully explained by the Bayesian/PP story – at the very most all we have done, Chalmers may insist, is to have explained the patterns of judgment that deliver the meta-hard puzzle. How might we make headway with this kind of apparent stalemate?

Chalmers (2018 fn 28) notes that in his (1990) he proposed a "coherence test"

for theories of consciousness, holding that the explanation of *reports* about consciousness must cohere with the explanation of consciousness itself. Here, we think our Bayesian story does especially well. For the various verbal reports (including the reports of puzzlement) flow from the same bedrock processing economy as do the simpler behaviors of other sentient life-forms. The brains of such animals would likewise infer mid-level latent variables capturing patterns in gustatory space, auditory space, visual space, and the various bodily patterns captured (when all goes well) by experiences of pain and pleasure. In all such cases, latent variables are inferred so as to deliver efficient (simple yet effective) means of selecting adaptive actions.

According to our story, the reports of qualitative states by beings such as ourselves reflect just these kinds of adaptively valuable grouping of patterns registered in the sensorium. Importantly, detailed PP accounts here show how interoceptive information (concerning our own bodily states) continuously impacts both exteroceptive perception and the selection of action, and how the self-prediction of our own patterns of reactions helps convince us that subjective states such as 'finding kittens cute' are as real as the kittens themselves (see Dennett (2015), Clark (2016b) (2017)).

Our distinctive capacities for puzzlement then arise because, courtesy of the depth and complexity of our generative model, we are able to see that these groupings (the redness of the objects, the cuteness of some animals) reflect highly certain information that nonetheless fails to fully mandate specific ways for the external world (or body) to be. We thus become aware that these states, known with great certainty, seem to belong to the 'appearance' side of an appearance/reality divide (see Allen (1997)).

It might then be asked why, on our story, the very same issues do not arise for beliefs (folk psychologically construed)? After all, we can also hold a belief with high certainty, and then imaginatively vary how the world might be consistent with that belief. I can strongly believe, as a referee suggested, that Milan is in Italy, but then imagine I am in a world where we have all been fooled into holding this belief. That's true. But nothing in our story implies that *whenever* such a pattern obtains, we will experience distinctive 'qualia'. Rather, the claim is that qualitative contents reflect mid-level sensory encodings apt for the selection of local action, and/or steeped in interoceptive information. These strikingly certain, sensorially-rich content states are then mistaken for something else (something 'beyond content') when we engage in certain kinds of imaginative exercise that hold them fixed while varying the distal realm. The fact that we can engage in similar exercises for less rich content-states shows only that we can manipulate them too. However, only the strongly sensory contents invite the familiar construct of 'qualia' onto the argumentative scene. This may be because the mid-level encodings track distinctive kinds of content, relative to which higher-level states of belief are somewhat thin and unidimensional, hence less likely to generate a colourful and vivid (note the very words we use) thought experiment. It may be, for example (Friston, Parr and deVries (2017)) that high-level beliefs reflect locations in discrete computational spaces, while the sensory encodings are defined over continuous spaces. In this kind of way, we would expect a mature science of the predictive mind to explain why we infer there to be a deep and critical difference between perception and belief.

Finally, what about the worry (David Chalmers, personal communication) that

our account targets not experience but certainty. To be sure, self-estimated certainty lies at the core of our story. Our claim is that when the brain estimates that a suite of mid-level re-codings, couched in terms of features such as redness, roundness, loudness, pulsatingness etc. etc., as highly certain, it can simultaneously compute that this vivid set of (perhaps 100% agent-certain) contents is consistent with multiple ways the real world might be - including no way at all, in the key cases of dreaming or hallucination. Creatures who self-estimate their own certainty in these ways, against the backdrop of a rich enough generative model, will infer that they are home to mysterious 'experiences' able to float surprisingly free of how things are in the world outside.. To further insist that the story needs to invoke a distinctive realm of 'experience' (rather than a realm of contents and probabilities) is simply to beg the question against any such account of the meta-hard puzzlement profile.

## 7. Conclusions

We think out story shows promise. It passes Chalmers 'coherence test' and accounts for the apparent differences between beliefs and percepts within a framework that is neither standardly 'qualia-realist' not standardly 'illusionist'. It is realist in that it identifies qualia with distinctive mid-level sensory states known with high systemic (and 100% agentive) certainty. But it is illusionist in that it depicts this complete certainty as itself a kind of illusion, plausibly induced to streamline action and choice, and defined over contents that, while distinctive, are not in any metaphysically salient way special.

What emerges is a picture of the paradigm conscious agent as a being who

scores rather well along three key – but potentially dissociable – dimensions. The first is the scope and depth (and especially the temporal depth - see e.g. Friston et al. (2017)) of the generative model of worldly states of affairs. The second (Seth (2013), Barrett (2017)) is the extent to which the use of that model is itself responsive to interoceptive information concerning the agent's own bodily states and self-predicted patterns of future reaction. The third – and the one we here identify as most important for the issues surrounding the meta-problem – is the capacity to keep inferred, highly certain mid-level sensory re-codings fixed while imaginatively varying top-level beliefs. This is what allows the advanced agent to understand that what she so clearly sees in the fountain just might turn out to be vodka (or even nothing at all), while remaining highly certain of the appearances themselves.

It is the presence of that puzzling capacity – itself realized by agentive control over precision assignments – that delivers the 'inference to qualia'. This occurs when, seeing that potential gap between this highly certain mid-level re-coding of the sensory evidence and our own top-level beliefs, a system infers the presence of a kind of mysterious intervening qualitative realm (Dennett's 'figment' perhaps) capable of strongly grounding while not quite necessitating beliefs about states of the distal world (or body). Our own qualitative experiences, this suggests, are not some kind of raw datum but are themselves the product of an unconscious (Bayesian) inference, reflecting the genuine (but entirely non-mysterious) combination of processes described above. Crucially, we do not infer that we have qualitative experiences *because* we see red, feel pain etc. Instead, the arrow of causality runs the other way. We see red because we infer a strangely certain and peculiarly independent dimension of 'looking red' as part of the mundane process of predicting the world.

We think what we have presented is the core of a substantive theory of conscious experience. But rather than argue for this, we claim only that the considerations on the table resolve the meta-hard problem. They explain why it is that some agents become puzzled (quasi-puzzled – recall Section 1) in the ways distinctive of debates concerning qualitative experience. Such agents are making inferences based on their capacities to use precision-weighting variations to deliver a grip on counterfactual scenarios in which appearance and reality come apart. They are then led to represent some of their representations (the highly certain mid-level encodings) as deeply special, opening the door to all the demons of the Cartesian mindset.

* Andy Clark and Sam Wilkinson were supported in part by EU ERC Advanced Grant XSPECT - DLV-692739. Karl Friston was funded by a Wellcome Trust Principal Research Fellowship (Ref: 088130/Z/09/Z). Thanks to Wolfgang Schwarz for useful discussion.

## References

Allen C (1997) Animal Cognition and Animal Minds in P. Machamer & M. Carrier (eds.) *Philosophy and the Sciences of the Mind* Pittsburgh University Press pp. 227-243.

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social cognitive and affective neuroscience*. 12(1), 1-23

Bayes, T. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*. **53**: 370–418.

Bergh, O. Van Den, Witthöft, M., Petersen, S., & Brown, R. J. (2017). Neuroscience and Biobehavioral Reviews Symptoms and the body : Taking the inferential leap. *Neuroscience and Biobehavioral Reviews*, *74*, 185–203.

Chalmers, D.J. (1990) Consciousness and cognition, [Online], http://consc.net/papers/c-and-c.html.

Chalmers, D.J. (1996) *The Conscious Mind*, New York: Oxford University Press.

Chalmers (2018) The Meta-Problem of Consciousness. *Journal of Consciousness Studies* 25, No. 9–10 pp. 6–61

Clark, A. (2000) A case where access implies qualia?, *Analysis*, 60 (1), pp. 30–37.

Clark, A (2013) Whatever Next? *Behavioral and Brain Sciences* 36: 3: p. 181-204

Clark, A (2016a) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind* (Oxford University Press, NY)

Clark, A. (2016b). Strange Inversions: Prediction and the Explanation of Conscious Experience. In B. Huebner (Ed.), *Engaging Daniel Dennett*. Oxford University Press

.

Clark, A (2017) Consciousness and the Predictive Brain, in K. Almqvist & A. Haag (eds) *The Return of Consciousness* (Stockholm: Axel and Margaret Ax:son Johnson Foundation) 59-74

Clark, A (2018) Beyond the 'Bayesian Blur': Probabilistic Brains and the Nature of Subjective Experience. *Journal of Consciousness Studies* Volume 25, Numbers 3-4, 2018, pp. 71-87(17)

Dennett, D. C. (1988) Quining Qualia. In: Marcel, A. & Bisiach, E. (eds.) *Consciousness in Modern Science*, Oxford University Press.

Dennett, D.C. (2015) Why and how does consciousness seem the way it seems?, in Metzinger, T. & Windt, J.M. (eds.) OpenMIND, Frankfurt am Main: MIND Group.

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience* 4:215. doi: 10.3389/fnhum.2010.00215

Frankish, K. (2016) Illusionism as a theory of consciousness, Journal of Con- sciousness Studies, 23 (11–12), pp. 11–39. Reprinted in Frankish, K. (ed.) (2017) Illusionism as a Theory of Consciousness, Exeter: Imprint Academic.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B* .29;360(1456):815-36.

. Friston, K. J., Parr, T. & de Vries, B. (2017) *Network Neuroscience*. The graphical brain: belief propagation and active inference. *Network Neuroscience*.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience* 6, 187–214. doi: 10.1080/17588928.2015.1020053

Friston, K., Rosch, R., Parr, T., Price, C., Bowman, H. (2017) Deep temporal models and active inference. *Neuroscience and biobehavioral reviews* 77:388-402.

Hawley, K. and Macpherson, F. (Eds.) (2011) *The Admissible Contents of Experience.* Wiley-Blackwell.

Hohwy, J (2013) *The Predictive Mind* (Oxford University press, NY)

Letheby, C., Gerrans, P. (2017). Self unbound: ego dissolution in psychedelic experience. *Neuroscience of Consciousness* 3:1-11.

Parr, T., and Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Sci. Rep.* 7:14678. doi: 10.1038/s41598-017-15249-0

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Schwarz, W. (2018) Imaginary foundations, *Ergo*, [Online], https://www.umsu.de/papers/imaginary.pdf.

Schwarz, W. (This Issue) From Sensor Variables to Phenomenal Facts

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in cognitive sciences*, *17*(11), 565-573.