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 set of behaviors and verbal reports that constitute the so-called 'hard problem of consciousness'. Chalmers (and many others) think that the meta-problem can, and perhaps should, be addressed independently of any substantive account of consciousness. We take an alternative approach. Using the tools of Bayesian psychology, along with considerations from a leading Bayesian process model ('Predictive Processing') we first provide a substantive (but revisionary) account of consciousness, and then argue that the resulting schema directly explains the meta-problem data. This, in turn, provides further evidence in favor of the substantive picture itself.

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 (this issue)) 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

[1] University of Edinburgh, UK
[2] Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London (UCL), London, United Kingdom
[3] University of Exeter, UK

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 (this issue) 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". He claims that the meta-problem is "neutral on the existence and nature of consciousness" and suggests that it is fruitful to address it prior to addressing – and while suspending judgment about – the hard problem itself.

We will take a different approach. We agree that everyone - whatever their view on the nature (perhaps even the very existence) of conscious experience - has a duty to explain the various behaviors of (apparent) puzzlement etc. that creatures like ourselves produce. But we will seek to link our own explanation of the puzzlements (puzzlement behaviors) directly to a substantive account of conscious experience itself. Nonetheless, the account we offer should be of interest even to those who remain skeptical about our attempt to grapple with the hard problem. For at the very least, we 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'.

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 that illusion, a disguised grip on our own patterns of response (e.g. to the substances we call 'sweet') becomes misleadingly reified as a kind of mysterious intervening state mediating between energetic stimulations at the sensory surfaces and behavioral outputs. In 'Quining Qulia' 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 themselves inferred causes, constructed to accommodate the raw sensory flux.

This 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 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 that best predict the evolving flux of energies across our sensory surfaces.

We first (very briefly) introduce ‘predictive processing’ as a general framework for understanding human cognition. We then outline that positive, albeit revisionary, story concerning the nature and origins of conscious experience, depicting conscious experience as the rolling product of a process of inference. We go on to show why agents thus constituted will, when a few other conditions are met, start to display puzzlement concerning their own states of consciousness. We end by considering some possible responses to the account on offer.

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 gradient-descent methods to steadily improve its guesses. Such a process results in the installment of a probabilistic model of

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

the distal causes (sometimes called 'hidden causes' or 'latent variables') that might be generating 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.

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). Intuitively, this corresponds to a kind of increasing abstraction as trained-up processing moves deeper and deeper into the system. For example, lower levels might learn about lines and edges, while higher levels learn about shapes, and still higher levels about persisting objects (for a simple demonstration, see Rao and Ballard (1999)). 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 that is simply the difference between current predictions and the sensory evidence. A coherent percept forms when a multi-level cascade of top-down guessing adequately accommodates the exteroceptive sensory data (from sight, sound, etc.).

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, and test, their own predictions (Friston, Rigoli et al (2015)). Living organizations are strongly driven by the need to maintain various forms of bodily and metabolic homeostasis (and allostasis). That means they need to anticipate changes that would threaten their bodily integrity, and take action in good time. For example, increasing body temperatures recruit counter-measures (such as sweating or moving to a cooler place) that seek to bring the system back within normal (predicted) bounds before real damage occurs. Our own bodily states, as tracked by inward-looking 'interoceptive' mechanisms, are *themselves* key targets for prediction and control (Seth 2013), Ashby (1947)).

It has recently been suggested (Clark (2017) – see also 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 organizational form that will use both interoceptive and exteroceptive sensory information to infer important features of its own body and world. Such creatures turn out (see e.g. Hohwy (2013)) to be broadly speaking 'Bayesian', insofar as they probabilistically combine prior 'beliefs' (the generative model) and new sensory evidence in ways apt to minimize prediction

[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). One of us has written elsewhere about this story, which fits well with the approach on offer. For that reason, we mostly omit further rehearsal here. See Clark (2017) and Clark (2016b).

errors and long-term sensory surprise. For the moment, however, we must assume that all this takes place in phenomenal darkness. 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 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. Knowledge of Semi-Opaque Mechanisms

Clark (2000) offered an account that was intended as a solution to the meta-problem, namely, an account of why conscious creatures might be puzzled by their own consciousness. At the core of that account was the observation (Chalmers (1996)) that creatures possessing a certain kind of genuine but merely partial knowledge of their own sensory processing mechanisms would be led to make just the kinds of inference concerning the directly-known presence, in their 'experience', of puzzling brute sensory qualities.

The argument went roughly like this. Imagine a language-enabled creature capable of distinguishing between (for example) otherwise identical red and yellow coffee mugs using visual means. Note that visual discrimination and language production both belong to the class of easy problems. Now imagine that that same creature enjoys some genuine, but limited, access to its own problem-solving strategies. Imagine that it can tell when it is using vision as opposed to, say, audition, smell, or touch, to solve a problem. In other words, it has access to facts about the sensory modalities involved in its own responses[6]. Now suppose that same creature is interrogated about its own successful mug-discrimination behavior. Knowing that it solved the puzzle by visual means, but not knowing any more about the processing involved, it would be forced (Clark claimed) to assert that "the mugs simply looked different". If pressed, it could note that the visually-detected difference was not shape or size based. It will soon be led to assert that the two mugs simply 'looked different' in respect of some other puzzling property X – one regularly associated with that mug, and with certain other items, but about which the creature cannot really say any more. For want of anything more to say, the creature calls that property 'color'. These kinds of simplifying models, reflecting real but partial access to our own processing, are the source (Clark argued) of our intuitions and puzzlements concerning phenomenal experience. Such a creature is well on the road to inventing an over-arching concept of 'qualia' – puzzling brute features – colors, shapes, sounds - apparently simply 'given' in its experience of the world.

Chalmers' (2018) response to that story is to doubt whether access to modality-specifying information is up to the task. His reasons are two-fold, one very general, and one more specific. The more specific point he raises is that:

[6] Clark (2000) made much of the idea that such knowledge was 'direct and non-inferential'. This remains true, insofar as it will still seem direct and non-inferential to the agent concerned. But that is compatible – as Clark noted in the original treatment - with the kind of sub-personal inference we highlight here.

> "in the case of belief we also have access to an attitude (believing rather than desiring, say), and it is not really clear why access to a modality as opposed to an attitude should make such a striking difference." Chalmers (2018) p.24

The more general point is that (from his perspective) all we have done *at best* is to explain a pattern of judgments (or quasi-judgments), not the existence of any actual qualitative states - see Chalmers (2018 p.9).

Chalmers categorizes Clark's (2000) approach as a version of what Frankish (2016)) calls 'illusionism'. Illusionists hold that a solution to the meta-problem (of explaining the various 'puzzlement behaviors') will solve or eliminate the hard problem itself. Clark has been wary of the full-strength eliminativist reading, according to which conscious experience is itself an illusion. Rather than eliminating consciousness and qualia, Clark sought to explain how and why they arise, while simultaneously showing that they are not quite what they seem. In what follows we revisit this project armed with the new scientific and conceptual tools introduced in section 2 above.

4. 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 dimension – an 'imaginary foundation'. Importantly, this foundation does not consist in the energetic readings themselves (the electrochemical signals registered at the peripheries) but will be a re-coding of those signals optimized to act as a kind of efficient go-between in the process of conditionalizing high-level beliefs upon new sensory evidence.

The best way to motivate this proposal is simply to note 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 the fountain – is water. It might be vodka instead. But we do seem to become very certain of something, where that something is, intuitively speaking, more like a bunch of phenomenal features than a high-level belief. Note that we should not say, for example, that it most surely appears to be water. For how it appears is itself answerable to all our background beliefs. To a differently-biased agent, it might appear *just as if* there is vodka in the fountain. What we need is something different, something cognitively slimmer, able to stand firm despite all that.

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. In so doing, they are poised to usher an appearance/reality distinction onto the cognitive stage. 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. 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.

What's most important for our purposes, however, is how such a process (of inference that includes imaginary foundations) might seem to the agent herself. Here, Schwarz makes a bold and pregnant conjecture. Imaginary propositions, he speculates, might correspond to the features that seem to populate phenomenal consciousness. In broad outline, the argument is easy enough. Imaginary foundations here act as a kind of cognitive go-between, re-coding electrochemical patterns in ways optimized for reasoning and action, while remaining *almost as certain* as the electrochemical perturbations themselves. That re-coding maintains maximal flexibility for top-level belief fixation, but plays the key epistemic role of presenting, in suitably compressed form, what was with greatest certainty established by the waves of sensory stimulation themselves. Much that seems puzzling about the nature and role of phenomenal features then falls neatly into place. In particular, Schwarz argues that it will start to seem (to that agent) as if the world includes dimensions of similarity and difference that are very securely known but that are not themselves fully determined by how the distal world actually is.

Schwarz' picture sheds new light on the argument from Clark (2000) rehearsed above. Recall that the mug-discriminating agent was forced to acknowledge that she was relying on *something*, where that something seemed not to consist in a fact about the world so much as a fact about experience. Schwarz' picture accounts for this and shows, in addition, why such an agent might feel especially certain of her phenomenal experience. But to do full justice to this proposal, we must now consider an additional, and 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, Kanai, Komura et al. 2015). 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. Furthermore, I will know (report) that I am fairly certain that this is a mug and, more importantly, I will have (a possibly subpersonal) belief I have absolutely no confidence in my beliefs about its color. Conversely, and returning to the case from Clark (2000), that perceiver will now report that she believes she is seeing two mugs, different in respect of some highly certain but otherwise mysterious property whose behavior in other real-world situations is captured by the communally handy concept of 'color'.

All this explains, we suggest, much of our agentive puzzlement concerning so-called 'qualia'. The agent knows these mid-level properties (compressed re-codings of electrochemical stimulations) with great confidence. But both the properties themselves and their degree of certainty (i.e., precision) are computed by entirely agent-opaque means. Crucially, 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 in that fountain, and realize it would look *just like this*.

At this point, intelligent systems 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. Crucially, 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 illusionism. If the term 'qualia' is constrained to pick out some kind of raw experiential data, then qualia are an illusions, 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 on a par with dogs, cats, and vicars. We return to this issue later in our treatment.

5. Making It Real

All this unpacks very gracefully in the modern setting of hierarchical Bayesian inference. The key move here is to appreciate what hierarchical inference brings to the table. Hierarchical inference is Bayesian belief updating under a hierarchical generative model. A hierarchical generative model – also known as empirical Bayes (Efron and Morris, 1973, Kass and Steffey, 1989) – implies that 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.

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 this 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".

Our intuition that there is a firm distinction between perception and belief now falls directly into place. What we think of as perceptual experience, this suggests, is nothing other than a set of abstract re-descriptions of the sensory evidence that are consistent with multiple interpretations of the kind that emerge as higher levels settle into a best-fit picture of how the distal world most probably is. Our ideas about perceptual experience thus reflect the fact that specific patterns of high mid-level certainty can be consistent with many distal causes. The idea that perception and cognition are distinct processes may have its roots (and limited grains of truth) exactly here. High mid-level sensory certainty constrains how we actually take the world to be. But advanced perceivers can deliberatively explore other ways the world might be, consistent with holding fixed those mid-level encodings.

The nature of the effort involved is (from a PP perspective) clear enough. It involves the deliberate control of the precisions assigned to various low and high-level beliefs. 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 is not possible. For them, their own 'imaginary 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.

6. 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 this kind of worry, asserting that despite the intellectual attractions of some form of illusionism "On my 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 that 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)).

Chalmers (2018) also asks why, on our kind of story, perceptual experiences but not beliefs acquire such apparently phenomenal feels? Part of the answer here may relate to the extent to which perceptual states are impacted by interoceptive information, giving them an unusually 'bodily' feel (see Seth (2013)). But additionally, we speculate that this intuition (of a qualitative difference between belief and perception) is itself a product of the process we have described. Because we are able counterfactually to vary our own precision-weighted processing, we can 'see' that the very same set of precise mid-level states could be consistent with different ways (reported at higher levels) the world actually is. The idea that there is a qualitative difference between perceiving and believing then emerges as just another unconscious inference rooted in our advanced capacities for deliberate precision-control and (hence) counterfactual reasoning.

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'. Instead, our intuitions about qualia, just like our intuitions about cats and dogs, are rooted in patterns of inference that attempt to bring the flux of sensory stimulations under an efficient multi-level predictive net.

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 and highly certain mid-level sensory states fixed while 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 rather than water, while remaining completely 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 belief, a system infers the presence of a kind of mysterious qualitative realm 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 dimension of 'looking red' as part of the mundane process of predicting the world.

In closing, we note that the staunch qualia realist may well resist our claim to have thereby sketched the broad shape of a solution to the hard problem even while allowing that progress has here been made with the meta-hard problem. The story on offer would then simply help 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.

* Andy Clark and Sam Wilkinson were supported in part by 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.

Ashby, W.R., (1947) Principles of the self-organizing dynamic system. *Journal of General Psychology*. 37:125-128.

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.

Cocchi, L., Sale, M.V., Gollo, L.L., Bell, P.T., Nguyen, V.T., Zalesky, A., Breakspear, M., Mattingley, J.B. (2016) A hierarchy of timescales explains distinct effects of local inhibition of primary visual cortex and frontal eye fields. Elife 5, e15252

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

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.

Efron, B., Morris, C. (1973) Stein's estimation rule and its competitors – an empirical Bayes approach. *Journal of the American Statistical Association* 68:117-130.

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

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

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., 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)

Kass, R., Steffey, D. (1989) Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association* 407:717-726.

Kanai, R., Komura, Y., Shipp, S., Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B* 370:20140169. doi: 10.1098/rstb.2014.0169

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

Murray, J.D., Bernacchia, A., Freedman, D.J., Romo, R., Wallis, J.D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., Wang, X.J. (2014) A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience* 17:1661-1663.

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) Explaining the phenomena

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