François Kammerer

The Hardest Aspect of the Illusion Problem — and How to Solve it

Abstract: *In 'Illusionism as a Theory of Consciousness', Frankish argues for illusionism: the thesis that phenomenal consciousness does not exist, but merely seems to exist. Illusionism, he says, 'replaces the hard problem with the illusion problem — the problem of explaining how the illusion of phenomenality arises and why it is so powerful'. The illusion of phenomenality is indeed quite powerful. In fact, it is much more powerful than any other illusion, in the sense that we face a very special and unique intuitive resistance when trying to accept that phenomenality is an illusion. This is bad news for illusionists, because this means that they cannot entirely model their explanation of the illusion of consciousness on the explanation of other illusions. Explaining this unique intuitive resistance to illusionism may therefore constitute the hardest aspect of the illusion problem. However, I think that this aspect of the problem is solvable. I will outline a possible solution, which is based on the hypothesis that our (illusory) introspective representations of phenomenal states characterize them as having unique epistemological properties and as playing a special epistemological role.*

In 'Illusionism as a Theory of Consciousness', Keith Frankish argues that phenomenal consciousness does not exist, but merely seems to exist. I will not try to criticize his arguments here, as I am quite convinced that some kind of illusionism must be true regarding phenomenal consciousness. What I want to do is to examine what I take to be the most difficult challenge for illusionist theories of consciousness,

Correspondence:
François Kammerer, Université Paris-Sorbonne, Paris, France.
Email: kammerer.francois@gmail.com

Journal of Consciousness Studies, **23**, No. 11–12, 2016, pp. 124–39

which does not receive any particular attention in Frankish's paper. I will explain why I think current illusionist theories of consciousness fail to meet this challenge, and I will outline a possible solution to this problem.

1. The hardest aspect of the illusion problem

Proponents of illusionism, says Keith Frankish, replace the hard problem of consciousness with the illusion problem (Frankish, this issue, p. 37). Instead of having to explain how and why phenomenal consciousness arises from the physical, they only have to explain why it *seems to us* that phenomenal consciousness exists, while it does not. They also have to explain why this illusion is so powerful.

While this problem may be much more tractable than the original hard problem of consciousness, I think that this problem has a 'hard' aspect which does not receive a lot of attention in Frankish's paper, and more generally in the work of proponents of illusionism. This aspect consists in the explanation of the following fact: the illusion of phenomenality is not only *very powerful*; it is, in fact, *much more* powerful than *any other illusion*. It is powerful in a very distinctive way, as we face a unique intuitive resistance when we try to accept the illusory nature of phenomenality.

Indeed, denying the existence of phenomenal states does seem crazy to many. Chalmers writes that in doing so one 'denies the evidence of our own experience. This is the sort of thing that can only be done by a philosopher' (Chalmers, 1996, p. 188). Searle writes about Dennett's illusionism that it 'denies the existence of the data'; and he comes close to calling this kind of position insane — 'surely no sane person could deny the existence of feelings' (Searle, 1997). Other examples of such statements abound.[1] And anyone can easily feel the strangeness of illusionism simply by focusing on her own experience. Let's say that you are having an experience of a red rose. Focus on your experience, and then try to think: 'I am not really having an experience endowed with the qualitative feel it seems to have. This is simply an illusion.' Chances are that you will find this idea deeply puzzling

[1] Bryan Frances writes, in a discussion of scepticism: 'I assume that eliminativism about feelings really is crazy' (Frances, 2008, p. 241). Galen Strawson writes that eliminativists who deny that there are mental states such that there is something it is like to be in these states 'do seem to be out of their minds', and that their position is 'crazy, in a distinctively philosophical way' (Strawson, 1994, p. 101).

and even hard to make sense of, even if you embrace illusionism from a theoretical point of view.

I think that this very strong intuitive resistance to illusionism has to be explained by illusionist theories. After all, the fact that we are so strongly reluctant to give up our belief in phenomenality characterizes the peculiar way in which we are subjected to the illusion of phenomenality. This intuitive resistance to illusionism is therefore a part of the *explanandum* of illusionist theories. Moreover, this strong intuitive resistance to illusionism, and the fact that it seems crazy, is probably the main reason why illusionism gained so little support amongst philosophers of mind, in spite of its numerous theoretical advantages. Explaining this strong intuitive resistance within an illusionist framework would therefore be useful for illusionists, as it would give them some dialectical leverage when faced with accusations of craziness. 'It is true that our theory seems crazy, but it explains very simply and naturally many things, including the very fact that it seems crazy.'[2] The problem is that, as I will try to show, explaining this strong intuitive resistance is not easy, but quite hard. It is, in fact, the hardest aspect of the illusion problem.

2. Why is this aspect the hardest aspect of the illusion problem?

Many proponents of illusionism, I think, would begin by saying that illusionist theories can *already* explain this intuitive resistance to illusionism in a rather natural way. They would draw a parallel with perceptual illusions. The illusion of phenomenality may arise because of some hardwired features of our introspective device, in such a way that this illusion has some degree of *cognitive impenetrability* (Frankish, this issue, p. 18). It persists even in the face of opposite beliefs. That could explain why, even if we are convinced illusionists, our intuition still pulls us away from illusionism. Similarly, in the Müller-Lyer illusion, we are still intuitively tempted to judge that the two lines have different lengths, even when we *believe* that they do not.

2 This is not to say that everyone would be fully satisfied by such an answer (Chalmers, 1996, pp. 188–9). However, I think that it is undeniable that a good explanation of the strong intuitive resistance to illusion would at least give *some* dialectical force to the illusionist position.

However, I think that the cases are very different on closer examination. The intuitive resistance we face when we try to think that consciousness is an illusion is special, *sui generis*. It cannot simply be explained by the fact that we have an ongoing, perceptual-like, cognitively impenetrable disposition to believe that we are conscious. Therefore, it cannot be explained simply by following the model of the explanation of perceptual illusions.

The difference amounts to this: what makes us reluctant to accept illusionism is not only that we are disposed to believe that we are conscious, it is also that we have difficulties *making sense of the hypothesis that we are not conscious while it seems to us that we are*. We struggle to simply *picture* what it would mean for consciousness to be an *illusion*. When I focus on my current experience of a red rose, I find the idea that I do not really have this precise feeling, but only an *illusion* of this feeling, not only *unlikely to be true*, but deeply weird, puzzling, and almost senseless.

Nothing similar happens in the case of perceptual illusions. Let's imagine, for example, that I am falling prey to the Müller-Lyer illusion, so that I believe that the two lines are as they seem: that they have different lengths (for example, because it is my first time looking at these lines and I have not measured them yet). Even in this situation, I would still easily *understand* what it would mean for me to be victim of an illusion in that case. The idea that the two lines are *not* as they seem, and that they have in fact the same length *in spite of appearances* would make perfect sense to me.

In order to make the contrast even starker, let's take the example of a perceptual belief I *do* hold. I *do* believe that I have two hands right now, on the basis of my perception of my two hands. My perceptual disposition to believe that I have two hands is very strong, and so is my corresponding belief. However, I have absolutely no problem *entertaining* and *understanding* the hypothesis that I may be a one-armed person victim of an hallucination. It would be *very hard* to convince me that it is the case, but the hypothesis, however far-fetched, is not at all difficult to *understand* or to *make sense of*.

This contrast shows that the intuitive resistance to illusionism regarding consciousness is pretty unique, and cannot simply be explained by the persistence of a very strong perceptual-like, cognitively impenetrable disposition to believe that we are conscious. The illusion of phenomenality is particularly powerful; it is much more powerful than other 'classical' perceptual illusions, and it is powerful in a very distinctive way.

This is a problem for illusionist theories of consciousness. Indeed, most recent illusionist theories of consciousness tried to model their explanation of the illusion of phenomenality on the explanation of other illusions. But if the illusion of phenomenality has a distinctive and special force when compared to other illusions, this means that this kind of explanation is doomed to fail. Indeed, this kind of explanation cannot account for the following contrast: we can easily represent to ourselves that a 'non-phenomenal' situation, which we grasp through perception for example, is illusory, while we struggle deeply to understand how a *phenomenal* situation, grasped through introspection, could be illusory.[3] That is why this aspect of the illusion problem — the explanation of the unique intuitive resistance we face when we try to accept illusionism — is the hardest aspect of the illusion problem.

3. Current illusionist theories of consciousness cannot solve the hardest aspect of the illusion problem: case studies

Nicholas Humphrey (Frankish, this issue, p. 17; Humphrey, 2011, chapter 2) states that the illusion of phenomenality is comparable to the visual illusion of impossible objects. Real physical objects, such as the 'Gregundrum' (Frankish, this issue, p. 17), when seen under a certain perspective, can give the illusion that we are facing an *impossible* object: a solid Penrose triangle. The same way, the introspective perspective that we take on some of our neural states is what gives us the illusion that we have *phenomenal sensations*, which are impossible objects (intrinsic, private entities).

The problem, from my point of view, is that we cannot explain the hardest aspect of the illusion problem simply by following this model. Indeed, when facing a Gregundrum, we do not have any particular intuitive resistance to entertaining the hypothesis that the apparent presence of a solid Penrose triangle could merely be an *illusion*. On the contrary, this hypothesis seems perfectly intelligible and sensible — actually, this may be the most natural hypothesis that comes to mind when facing what seems to be an *impossible* object, such as a

[3] Here I simply want to emphasize the existence of such a contrast. I will give my own explanation of why this contrast exists in Section 5.

Penrose triangle. So, this model cannot explain the peculiarity of the illusion of phenomenality.

Similarly, Michael Graziano has an illusionist theory of consciousness, the 'attention schema theory', that states that the brain monitors its own internal attentional processes in a schematic and fallacious way. Instead of representing what is really there — a set of complex attentional processes — it represents itself as being in states of awareness, where awareness is depicted as a 'fluidic substance', an 'experience', a 'sentience' (Graziano, 2013, p. 80); hence the fallacious impression that we are in qualitative and primitive *phenomenal states*. And, because this impression arises from some features of the very architecture of the brain, any theory that goes against this impression is judged very counter-intuitive. Quite in the same way, Graziano explains, Newton's theory of light, according to which white light is not pure and simple, but composed of all the other coloured lights, struck everyone as highly counter-intuitive when first formulated, because the human visual system is hardwired to depict white light as being simple, pure, and primitive.

However, again, such a model cannot account for the unique intuitive resistance we face in the case of illusionism regarding consciousness. Indeed, in the case of Newton, even though people were strongly disposed to believe that white light was pure, simple, and not composed, I do not think they had any difficulty *understanding* what it would mean for white light to be *profoundly different* from what it seemed to be. However surprising and weird this theory must have seemed to them, I do not think they had trouble picturing what Newton's theory meant: that white light *is not at all like our perception of it depicts it*. This thesis may be counter-intuitive, but it is not difficult to make sense of. But, on the other hand, we *do* have trouble fully grasping what it means for consciousness to be an illusion; we *do* struggle to make sense of this idea. Therefore, Graziano's account cannot solve the hardest aspect of the illusion problem.

Derk Pereboom's illusionist theory of consciousness (Frankish, this issue, pp. 17–8; Pereboom, 2011) is quite similar to the theories I just discussed. He suggests that introspection misrepresents phenomenal properties by depicting them as having a qualitative nature that they do not have, quite in the same way sensation represents external objects as having primitive colour properties ('edenic colours' — Chalmers, 2006; Pereboom, 2011) that they do not have. This hypothesis alone is not able to solve the hardest aspect of the illusion

problem, for the same reasons I already mentioned. Indeed, if this hypothesis were true, why would we have such difficulty making sense of illusionism regarding consciousness, while illusionism regarding primitive colours seems rather unproblematic? Indeed, the thesis according to which external objects only *appear* to bear primitive colours, but do not really bear them, is quite widely accepted since Galileo and Descartes (and has been on the philosophical market at least since Epicurus). And this thesis does not seem to be particularly difficult to *understand* or to make sense of.

Pereboom then makes an additional hypothesis: maybe illusionism about consciousness seems much less credible than about secondary properties such as colours because 'we cannot check the accuracy of introspection, as we can that of perception, by adopting different vantage points, using measurement instruments, and so forth' (Frankish, this issue, p. 18; Pereboom, 2011, p. 23). However, this hypothesis does not help much. After all, a physicist could build a detector D, to detect a new particle Z, and for a little while she may very well own only *one* copy of this detector, and have no other way to detect Z. The detections of Z made with D could therefore not be checked for accuracy using any other means. Would that mean that our physicist would have huge trouble *making sense* of the hypothesis that D is in fact inaccurate? Would she find such an hypothesis extremely puzzling, and barely intelligible? This seems quite unlikely. But this shows that the inability to check the accuracy of a given source of information cannot explain why subjects find the idea that the information provided by the source is illusory so puzzling and difficult to accept.

4. The theoretical introspection hypothesis

Current illusionist theories of consciousness cannot solve the hardest aspect of the illusion problem, mostly because they try to model their explanation of the illusion of phenomenality on the explanation of other 'classical', perceptual illusions. But the theories resulting from such a modelling cannot do the job. Indeed, the illusion of phenomenality has a distinctive and unique force *when compared with other illusions* — and this cries out for explanation.

I now want to outline how we could solve this difficulty. First, we have to acknowledge the particular force of the illusion of phenomenality, and the unique predicament we find ourselves in when we try to accept that consciousness is an illusion. By acknowledging this

difficulty, we do not weaken the illusionist position — quite the contrary, in my opinion.

Second, we have to develop an hypothesis that could account for the unique intuitive resistance to illusionism about consciousness; for example, an hypothesis concerning introspective representations and introspective processes, which could explain why we are so resistant to illusionism. I will now present an hypothesis I think could do the job. Call it the 'Theoretical Introspection Hypothesis' (TIH).[4]

Theoretical introspection hypothesis: suppose that, as proponents of the 'theory-theory of self-awareness'[5] say, introspection (including in this case phenomenal introspection) is a theoretically informed activity, where the relevant theories are our naïve theory of mind and our naïve epistemology. Suppose that these theories are innate and modular, that is to say, informationally encapsulated (though they are informationally related to each other — they may for example be implemented in the same module). Introspective representations of phenomenal states are deeply theoretically informed by these theories.[6]

The TIH states that the content of these theories includes, so to speak, the following statements:

(1) Minds can take up information about states of affairs, and then can use that information to form beliefs about them: states of affairs *appear* to minds.
(2) The way minds do that is that they are *affected* in a certain way, they have certain *experiences*.
(3) The properties of experiences determine what *appears* to the mind, and a state of affairs appears to the mind in virtue of these properties of experiences. For example: an experience of a red circle is an affection of the mind in virtue of which the presence of a red circle appears to the mind.
(4) Take all the cases in which a certain state of affairs A appears *veridically* to a subject S, and consider what all these cases have in common regarding the way in which S is affected. What they have in common is a state E, which is an *experience of A*.

4 For a first suggestion of an hypothesis of this kind, in a different vocabulary and within a different framework, see Kammerer (2016).

5 For a definition and a critical overview of the theory-theory of self-awareness, see Nichols and Stich (2003, pp. 164–9).

6 Frankish quickly mentions this kind of view (Frankish, this issue, p. 36).

Something is a part of E *if and only if* this thing is part of the way in which S is affected in *all the cases* in which A appears veridically to a subject S.[7]

(5) Appearances can be fallacious, and a mind can be deceived by the way states of affairs appear. And here is what happens in cases of fallacious appearances: when a subject S has a fallacious appearance of A, S is affected *in exactly the same way* as in cases of veridical appearances of A, except that A is not the case. That is to say, when a subject S has a fallacious appearance of A, it is in state E (E being, and being nothing but, what is common to the way S is affected in all the cases in which A appears veridically to her), but A is not the case.[8]

Of course, our naïve theory of mind/epistemology does not literally *contain* these statements. Naïve theories, when we conceive of them

7 One could then object that according to this thesis there would be states of affairs such that there can be no experiences of them according to our theory of mind, because there is no common element to the way in which subjects are affected when these states of affairs veridically appear to them, while we still commonly and naturally speak of *experiences* of these states of affairs. Let's take, for example, *the presence of a car*. The presence of a car can appear to me through very different affections (for example, what we call 'visual experiences', 'auditory experiences', 'olfactory experiences'), and there is probably nothing common, from an experiential point of view, to all these affections. However, we can still speak of an 'experience of a car'. That seems to contradict the idea that our naïve theory of mind does not allow for such a thing. My answer is the following: 'being an experience of a car' is indeed not an experiential kind *stricto sensu*, but is simply a secondarily constructed kind, which denotes a disjunction of very numerous (in fact, potentially infinite) and distinct 'real' experiential kinds, such as: 'being a visual experience of a white car-shaped thing of that size, in this perspective, at such distance', 'being an olfactory experience of oil, gasoline, with a soupçon of hot metal', etc. Certainly, for each of these 'real' experiential kinds, there is something that is common to the way in which the subjects are affected in all the situations in which the relevant states of affairs veridically appear to them, and in virtue of which these states of affairs appear to them.

8 So, according to our naïve theory of mind/epistemology, minds can be deceived by appearances because, in cases of fallacious appearances, they are affected *in exactly the same way* as in cases of veridical appearances. So, what is common, in the overall experiential state of a subject, in all cases of veridical appearances of something to a subject, must also be present in the overall experiential state of a subject undergoing a fallacious appearance of the same thing. What explains the necessity of (5), from the point of view of our naïve theory of mind/epistemology, is, roughly, the idea that if a mind was affected in a *different way* in a case of fallacious and veridical appearances of the same thing, it would be in a position to distinguish, on the sole basis of the appearances, between fallacious and veridical appearances of the same thing; but then 'deceptive appearances' could not exist, at least not in the full and simple sense used by our naïve theory of mind/epistemology.

as innate and modular, do not really consist in statements, but in sets of capacities and inferential mechanisms. However, our naïve theory of mind/epistemology could very well be a series of capacities and inferential mechanisms which function in such a way that it would correspond to these statements. Actually, I think that it is rather plausible, given that these statements themselves seem intuitively quite plausible.[9,10]

According to the TIH, phenomenal introspection is deeply theoretically informed by these theories. Therefore, when we introspect phenomenal states, we represent ourselves as being in states that are *experiences*, in the sense described by the theory: states in virtue of which certain states of affairs appear to us, which are gifted with special mental and epistemological properties, so that they satisfy the definition of *experiences* according to the theory — they are represented as satisfying (1–5), in virtue of the functional/inferential role of the relevant representations.[11]

5. How the TIH can solve the hardest aspect of the illusion problem

Let us suppose that the TIH is true, and that our introspective processes and representations are theoretically determined in the way I just described. Now, let us ask the question: what happens when we think intuitively — that is to say, within the framework of our naïve theory of mind/epistemology — that someone is undergoing an

9 This is not to say, of course, that every philosophical theory endorses these statements. Disjunctivists, for example, notably reject (5), and (4) could probably be questioned by some contemporary theories of consciousness.

10 It should be noted that there is a 'naïve' and intuitive concept of *appearance* that does not correspond to the concept of appearance that I just described as being the one used by our naïve theory of mind/epistemology. For example, we have a purely epistemic concept of appearance, such as the one we express linguistically when we say 'Einstein's theory of relativity appears to be true'. I think that this concept can be ultimately defined by our naïve theory of mind/epistemology in terms of phenomenal appearances and beliefs, but I suggest simply setting aside this issue, which is beyond the scope of this paper.

11 It should be noted that, from my point of view, given that (1–5) determine the content of our concepts of phenomenal states, anyone who claims that these five points do not characterize accurately our mental functioning cannot be a full-blown phenomenal realist, and is necessarily committed (though maybe only implicitly) to a form of illusionism. Of course, given that this commitment would be merely implicit, we should certainly not expect every philosopher who denies that the functioning of our mind satisfies (1–5) to explicitly endorse illusionism.

illusion? We think that this person has a fallacious appearance of something. In most cases, this creates no problem: there can be fallacious appearances of many things, and therefore many things can be merely *illusory*, even though we don't necessarily *believe* that they are illusory. For example, let us take the example of an appearance of a red rose. We think that a red rose (or, more precisely, the presence of a red rose) can appear to a subject. Let us take all the cases in which a red rose appears veridically to a subject, and consider what is common to all these cases, regarding the way in which the subject is affected: this common element is an *experience* of a red rose. And when we think that someone is undergoing an *illusion* of a red rose, we think that this person is having an experience of a red rose (she is affected in exactly the same way as she is affected in all the cases of veridical appearances of a red rose), except that there is no red rose. So far, so good.

However, in the case of consciousness, and *only* in the case of consciousness, something troubling happens. When we try to think intuitively — using the mental resources of our naïve theory of mind/epistemology — that our current experience of a red rose, for example, is purely illusory, and that we in fact have no experience of a red rose but merely a *fallacious appearance* of an experience of a red rose, we encounter a peculiar difficulty. In order to see why, let us consider what is common, regarding the way in which subjects are affected, to all the cases of *veridical appearances* of an *experience* of a red rose. This common element would be an *experience* of an experience of a red rose, so to speak — a second-order experience. But, given the principles of our naïve theory of mind/epistemology, this state will crucially *include* the first-order experience it is an experience of[12] — it will include an experience of a red rose. Indeed, an experience of a red rose is *part* of what is common, regarding the way in which subjects are affected, to all cases of *veridical appearances of an experience of a red rose*.[13] But now what happens when

[12] Many philosophers argue that experiences of experiences not only *include* first-order experiences, but are actually *nothing but first-order experiences*, so that there are *no second-order experiences* distinct from first-order experiences. This is sometimes remarked by people who insist that there is no specific 'phenomenology of introspection' (Lycan, 1996; Shoemaker, 1994; Siewert, 2012). I will not say more about this kind of thesis here, as it is beyond the scope of this paper.

[13] Indeed, in all the situations in which it appears veridically to a subject that she has an experience of a red rose, (1) she will have an experience of a red rose (by definition of

we try to think intuitively that we have a fallacious appearance of an experience of a red rose? We try to think that we are in a state which is an experience of an experience of a red rose, without having an experience of a red rose. But this is impossible, as the experience of an experience of a red rose has to include an experience of a red rose.

Let's sum that up: when we try to think intuitively that we have a fallacious appearance of a given experience, we try to think that we are affected *in exactly the same way* as we are affected *in all the cases in which it veridically appears to us* that we have this given experience. But, in all these cases of veridical appearance, the way in which we are affected crucially includes the fact that we have this given experience. So, we cannot be affected in a way that is exactly similar to cases of veridical appearances of the experience *without* having this given experience — that would lead to an obvious contradiction, given the content of the representations provided by our naïve theory of mind/epistemology.

So, what the TIH predicts is that, given the nature of the naïve theory of mind/epistemology that determines our introspective representations, we are unable to think of second-order experiences that do not *include* the first-order experiences they are experiences of. Therefore, we are unable to think of a fallacious appearance of an experience. This peculiar difficulty arises in the case of consciousness, and *only* in the case of consciousness. For any other thing, we have no difficulty thinking of a *fallacious appearance* of this thing.

If the TIH is true, then, it means that, to the extent that we use an *intuitive* and *naïve* concept of illusion (fallacious appearance), an '*illusory experience*' is a contradiction. This explains the unique intuitive resistance we face when we try to accept illusionism, given that it explains why we cannot fully make sense of the idea that our experiences are purely illusory, while we don't have this problem in the case of any other thing.

Of course, in this view, illusionism is still a perfectly *coherent* hypothesis. But we need to use another concept of illusion to make it

the situation, which is a situation of *veridical* appearance of such an experience); (2) this experience of a red rose will be part of the way in which subjects are affected (by definition of what an experience is according to our naïve theory of mind/epistemology). So, if we consider what is common, regarding the way in which a subject is affected, to all the situations in which it appears veridically to the subject that she has an experience of a red rose, we will find that this common element necessarily includes an experience of a red rose.

completely intelligible — not the naïve concept of illusion, provided by our naïve theory of mind/epistemology, but rather a functional concept of illusion. For example, we have to use a concept of illusion that states that an illusion simply consists in, say, a state that grounds an ongoing, systematic, and cognitively impenetrable disposition to believe something false. When we use such a concept of illusion, illusionism regarding consciousness is not problematic at all, and it is quite easy to understand and to make sense of.[14]

The TIH therefore explains why denying the existence of genuine phenomenal states is 'the sort of thing that can only be done by a philosopher' (Chalmers, 1996, p. 188): we can coherently think that consciousness is an illusion only if we use a highly abstract and theoretical concept of illusion. But, as soon as we try to intuitively make sense of this idea, we are back using our naïve and innate concept of illusion, and we cannot help but find illusionism about consciousness deeply puzzling and barely intelligible. The moment we turn away from the abstract, objective, and impersonal understanding of illusionism about consciousness (in which the illusion of phenomenality is defined in purely functional terms), and try to consider what illusionism 'amounts to' or 'means' in an intuitive manner, we find this position to be unacceptable, preposterous, and hardly intelligible. However, it is crucial to note that this obviously does *not* prevent illusionism (correctly understood) from being true.

The TIH also explains why one very common objection to illusionism regarding consciousness relies on the idea that, in the case of consciousness, there is no appearance–reality gap, and that therefore there can be no fallacious appearance of consciousness. Frankish (this issue, p. 32) is right to note that, in order for such an objection to succeed, it has to presuppose a conception of *appearances* that begs the question against illusionism. However, the TIH explains why this objection

14 It is also interesting to note that, according to my view, the thesis according to which *phenomenal states* do not exist, taken in itself (and if we don't add the idea that these states nevertheless *appear* to exist), is also not problematic at all — we find it very understandable and intelligible, even if it contradicts our beliefs. This is quite obviously the case, as most of us, for example, have no difficulty conceiving of a possible world in which there are no phenomenal states at all. What is problematic is the idea that phenomenal states do not exist *even if they appear to us to exist* (i.e. given our epistemic situation). This last idea, according to my view, is deeply problematic in virtue of the fact that the content of our theory of mind makes it so that we intuitively conceive of what appears to us in a way that implies that phenomenal states cannot appear fallaciously.

nevertheless has some intuitive weight: it is because this question-begging conception of *appearances* happens to be our naïve, innate, and intuitive conception of appearances, provided by our naïve theory of mind/epistemology.

Overall, I think that the TIH is able to give a solution to the hardest aspect of the illusion problem. The fact that illusionism about consciousness is intuitively unacceptable and almost senseless is a by-product of the simplistic structure of our naïve theory of mind/epistemology, and of the manner in which this theory deeply determines both the content of our introspective representations of phenomenal states and the way we intuitively conceive of fallacious appearances of things. Illusionism about consciousness, in order to be acceptable, has to be understood while using a highly abstract and theoretical conception of 'illusions' and 'appearances'. As soon as we seek an *intuitive* grasp of illusionism, this position strikes us as absurd.[15]

6. Conclusion

I think that the hardest aspect of the illusion problem — i.e. the explanation of the unique intuitive resistance we experience when facing illusionism about consciousness — cannot be solved if we understand the origin of the illusion of phenomenality simply on the model of other classical, perceptual illusions. This is why current illusionist theories of consciousness fail to solve this problem. However, I think that it is possible to solve it, if we recognize the peculiarity of introspective representations of phenomenal states. The TIH is an hypothesis that attempts to do so.

To speak broadly and metaphorically, this hypothesis states the following. Our naïve theory of mind/epistemology provides a mental tool to think about appearances (and potentially deceiving

[15] I claim that the TIH can solve the hardest aspect of the illusion problem. One could then wonder if it is sufficient to solve the full *illusion problem* by itself, or if we have to use additional claims regarding phenomenal introspection. Can we explain why it seems to us that we are phenomenally conscious even though we are not, simply by positing that our introspective representations have the kind of content they have according to the TIH? If we say 'yes', we claim that we can explain all the peculiar features that introspection presents phenomenal states as having (features that nothing has in reality) simply by the fact that our introspective representations have the content described by the TIH. I think that it is in fact possible to give such an explanation, but it would take a long argument to make a proper case for that claim. I therefore choose to set aside this more general issue, and to focus merely on the hardest aspect of the illusion problem.

appearances). This mental tool consists in our capacity to represent *experiences*, as ways in which minds are affected. However, this mental tool happens to be quite rough and rudimentary. Because of the rough representational structure of this mental tool, a problem arises when we try to apply this mental tool to itself, so to speak — that is to say, when we try to distinguish between *experiences* and *appearances of experiences that do not include these experiences*. Such a distinction is impossible, because of the simplistic structure of this mental tool. Therefore, we end up being incapable of thinking intuitively of a *fallacious appearance of a given experience*. This explains why illusionism about consciousness can never be intuitively acceptable or even intuitively fully intelligible, while we have no difficulty making intuitive sense of illusionism regarding *any other entity*. Of course, the TIH implies that our naïve theory of mind/epistemology is rough and simplistic, but this should not come out as a surprise. After all, all our other naïve theories are simplistic, and widely inaccurate.[16] In fact, it would be rather surprising if one of our naïve, innate theories happened to be perfectly accurate and satisfying.

The TIH may have the details of the story wrong, but I think that the main idea behind the hypothesis is correct: if we want to account for the unique strength of the illusion of phenomenality, we have to understand that introspection represents phenomenal states in a way that characterizes them as having special epistemological properties, and that in turn accounts for the difficulty we face when we try to give the 'illusion treatment' to phenomenal states — when we try to think of them as purely illusory. My guess is that a fully satisfying illusionist theory of consciousness would have to say something in the vicinity in order to solve the hardest aspect of the illusion problem.[17]

References

Chalmers, D.J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*, New York: Oxford University Press.

16 Our naïve physics, for example, is full or errors and approximations concerning the features of physical objects. It notably contains the idea that there are basic forces such as 'sucking' (used, for example, in a vacuum cleaner), of that the movement of bodies is caused by a kind of 'impetus', etc. (Hayes, 1978; McCloskey, 1983). The same could be said of our naïve biology.

17 I would like to thank Samuel Webb, Fabien Mikol, Keith Frankish, David Chalmers, and two anonymous referees for their comments and their help.

Chalmers, D.J. (2006) Perception and the fall from Eden, in Szabo, T. & Hawthorne, J. (eds.) *Perceptual Experience*, pp. 49–125, Oxford: Oxford University Press.

Frances, B. (2008) Live skeptical hypotheses, in Greco, J. (ed.) *The Oxford Handbook of Skepticism*, pp. 225–244, Oxford: Oxford University Press.

Frankish, K. (this issue) Illusionism as a theory of consciousness, *Journal of Consciousness Studies*, **23** (11–12).

Graziano, M. (2013) *Consciousness and the Social Brain*, Oxford: Oxford University Press.

Hayes, P. (1978) The naive physics manifesto, in Michie, D. (ed.) *Expert Systems in the Micro-Electronic Age*, Edinburgh: Edinburgh University Press.

Humphrey, N. (2011) *Soul Dust: The Magic of Consciousness*, Princeton, NJ: Princeton University Press.

Kammerer, F. (2016) Conscious experiences as ultimate seemings: Renewing the phenomenal concept strategy, *Argumenta*, **1** (2), pp. 233–243.

Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.

McCloskey, M. (1983) Naïve theories of motion, in Gentner, D. & Stevens, A. (eds.) *Mental Models*, Mahwah, NJ: Lawrence Erlbaum Associates.

Nichols, S. & Stich, S. (2003) How to read your own mind: A cognitive theory of self-consciousness, in Smith, Q. & Jokic, A. (eds.) *Consciousness: New Philosophical Perspectives*, Oxford: Oxford University Press.

Pereboom, D. (2011) *Consciousness and the Prospects of Physicalism*, Oxford: Oxford University Press.

Searle, J. (1997) *The Mystery of Consciousness*, New York: The New York Review of Books.

Shoemaker, S. (1994) The phenomenal character of experience, *Philosophy and Phenomenological Research*, **54** (2), pp. 291–314.

Siewert, C. (2012) On the phenomenology of introspection, in Smithies, D. & Stoljar, D. (ed.) *Introspection and Consciousness*, Oxford: Oxford University Press.

Strawson, G. (1994) *Mental Reality*, Cambridge, MA: MIT Press.