

# Awareness and Equilibrium

Brian Hill\*

August 21, 2010

## Abstract

There has been a recent surge of interest among economists in developing models of doxastic states that can account for some aspects of human cognitive limitations that are ignored by standard formal models, such as awareness. Epistemologists purport to have a principled reason for ignoring the question of awareness: under the equilibrium conception of doxastic states they favour, a doxastic state comprises the doxastic commitments an agent would recognise were he fully aware, so the question of awareness plays no role. The objective of this paper is to scrutinize this argument. A thesis underlying the argument, which we call the *independence of doxastic commitments with respect to awareness*, is identified, and examples are given where it appears to be violated. By considering these examples, one can get an idea of the price of accepting this thesis. On the one hand, one can escape the conclusion that the thesis is violated, but only at the expense of another principle espoused by all major formal models of belief, which we call *constant doxastic rest*; and abandoning this principle necessitates extensive revision of current models of belief. On the other hand, there are epistemologically valid reasons for thinking that the thesis fails to hold in the examples, which have to be rebutted if the thesis, and the equilibrium justification for ignoring the issue of awareness, are to be retained.

**Keywords** Bounded rationality; awareness; doxastic states; cognitive equilibrium; belief change; formal epistemology.

The main formal models of states of belief, or *doxastic states*, which have been proposed and used by epistemologists, as well the classic models of doxastic states employed by economists, share a central assumption. Whether one represents doxastic states by consistent sets of sentences closed under logical consequence, or equivalently, sets of possible worlds (Hintikka, 1962; Aumann, 1976;

---

\*HEC Paris. 1 rue de la Libération, 78351 Jouy-en-Josas, France. [hill@hec.fr](mailto:hill@hec.fr).  
[www.hec.fr/hill](http://www.hec.fr/hill).

Stalnaker, 1984; Gärdenfors, 1988), by probability measures or sets of probability measures (Ramsey, 1931; Savage, 1954; Jeffrey, 1972), or by combinations of sets of sentences and sets of probability measures (Levi, 1980), to take but a few important examples, one assumes that an agent's doxastic state at a particular moment can be represented by a (logically or probabilistically) well-behaved "structure" on the set of sentences of a given language.<sup>1</sup>

Of course, such an assumption has long been identified as problematic, if the purpose of these models is to describe the actual states of belief of agents in a "non-idealised" way. In particular, it implies that these models cannot take into account the agent's state of awareness (Dekel et al., 1998). However, the fact that an agent is aware or unaware of particular issues does seem to play a significant role in his decision-making, and that of others. Accordingly, there have been several recent efforts, particularly among economists conscious of the potential importance of the notion of awareness in strategic situations, to develop models which can capture awareness, and its consequences for behaviour (Heifetz et al., 2006; Halpern and Rêgo, 2007).

To date, epistemologists have not been impressed with such alleged problems with the models they use, and have made little effort to incorporate awareness into their models. A major reason for this is the particular conception many of them have of what their models are intended to represent: not the agent's "raw" beliefs, if there is such a thing, but the beliefs the agent would have were he allowed to come to a state of cognitive equilibrium. In any such equilibrium, the agent is fully aware; by considering only such equilibria, the epistemologist can ignore the question of awareness.

The objective of this paper is to scrutinize more closely this defence of the standard formal models used in epistemology, with an eye to assessing the price of ignoring the issue of awareness. We first identify an assumption underlying the equilibrium argument for these models, namely the thesis of independence of doxastic commitments with respect to awareness (Section 1). In Section 2, we consider examples which are *prima facie* violations of this thesis. These examples do not prove that the thesis of independence of doxastic commitments with respect to awareness is false, for one can find analyses which do not require it to be violated. However, such analyses violate another principle, which is shared by virtually all current models of doxastic states, that we call constant doxastic rest. In Section 3, we consider some unwelcome consequences of rejecting constant doxastic rest: they emphasise that the equilibrium conception of doxastic states does not come for "free", but may require some important modifications in current models and theories of belief. In Section 4, we consider arguments

---

<sup>1</sup>For the purposes of this paper, we can set aside the question of the nature of the objects of belief, and assume them to be sentences.

for rejecting the thesis of independence of doxastic commitments with respect to awareness; these arguments suggest that, contrary to the generally accepted view, incorporation of awareness into models of belief may be justified on epistemological grounds.

## 1 Cognitive Equilibrium and Awareness Change

It is well-known that the main formal models of belief used by economists and epistemologists do not do justice to human cognitive limitations. These models represent doxastic states as consistent, and such that all (logical or probabilistic) consequences of beliefs are drawn; ordinary human beings just do not seem to live up to these standards of consistency and completeness. As mentioned above, among philosophers, proponents of the models explicitly claim that they have a principled reason for ignoring such problems. A clear and eloquent defence of this claim, which is representative of the consensus in the field, is provided by Isaac Levi.<sup>2</sup>

I suggest that we distinguish between the standard for serious possibility to which X is *committed* at time *t* and X's *awareness* at *t* of the standard to which he is committed; or equivalently of the corpus to which he is committed. . . .

Thus, changes in X's awareness of his commitments at *t* ought to be distinguished from changes in X's commitments. The former sort of change may be compared to a shift towards an equilibrium. The more fully aware X is of his commitments, the closer he is to a state of cognitive equilibrium. On this analogy, X is committed at *t* to a state of cognitive equilibrium whether he has actually attained it or not. The features of rational equilibrium I have been discussing have been introduced by an appeal to those functions which X's corpus of knowledge ought ideally to perform.

The account of the revision of knowledge . . . I am aiming to construct here prescribes shifts from one state of cognitive equilibrium to another without prescribing details of the psychological or social changes which are made in implementing the revision. (Levi, 1980, pp10-11)

---

<sup>2</sup>See Levi (1991, Ch 2), Gärdenfors (1988, §1.2) and Bradley (2007, §4.1), to mention but several examples, for other formulations of similar positions, albeit in slightly different terms. Although there are subtle differences among the notions of cognitive equilibrium proposed in the literature, the general points extracted below are common to all, so the arguments proposed in this paper, although they are couched in the terms used by Levi (1980), apply irrespective of the notion of cognitive equilibrium adopted.

Epistemologists who propose and use the sorts of formal models mentioned above are interested in the beliefs the agent is committed to – his *doxastic commitments* – rather than those he recognises himself as having: those he would acknowledge were he fully aware, rather than those he does identify in his state of partial awareness. As is clear from the quote, this delimitation is supposed to allow one to ignore the problem of human cognitive limitations. One can continue to model doxastic states as fully consistent and closed under consequence since the doxastic states of interest to epistemologists are merely “idealised” cognitive equilibria. Moreover, the question of change need only be considered in terms of such ideal states, for only the move from one cognitive equilibrium – or closed and consistent set of doxastic commitments – to another is of interest. Under this *equilibrium conception* of the study of beliefs and belief change, awareness is an epiphenomenon, and as such can be ignored.

This conception tacitly rests on what may be called the thesis of the *independence of doxastic commitments with respect to awareness* (IDCA):

**(IDCA)** Changes in awareness alone do not affect the agent’s doxastic commitments.

Suppose that at moment  $t$  John is doxastically committed to sentence X. According to these models, the belief that X is in his doxastic state at  $t$ ; the doxastic commitment to X is in the cognitive equilibrium that is associated with him at  $t$  (and that he would be in were he fully aware). Suppose however that at  $t$  he is not aware of this commitment, because he is not aware of the issues involved (he is not aware of the sentence X itself). One step in the ‘shift towards an equilibrium’ described by Levi is becoming aware of X. However, if, on merely becoming aware of the issue, John changes his beliefs – and his doxastic commitments – and ceases to believe that X, then the cognitive equilibrium associated with him – his doxastic state, as modelled by the theories mentioned above – changes. If such changes in belief can occur rationally, the equilibrium conception is severely weakened.

First of all, the contention that there is an equilibrium is thrown into doubt. A change in awareness is supposed to be a shift towards equilibrium, but a shift towards equilibrium should not change the equilibrium towards which one is shifting: if it does, one may justifiably wonder whether one will ever reach an equilibrium at all. One might reply by refining the concept of cognitive equilibrium so that, by definition, it consists of the doxastic commitments the agent would have on becoming aware of all sentences, whether or not these include his current doxastic commitments. But then the possibility of changes in doxastic commitments on mere change of awareness threatens the consistency of this definition. If such changes are possible, then there is a risk that the order in which one becomes aware may affect the doxastic commitments one holds when one attains

full awareness. Of course, if the order of awareness change has such effects, this notion of cognitive equilibrium is not well-defined: there may be no unique set of doxastic commitments which the agent would have were he fully aware, for such a set would depend on the details of how he attained this awareness.

Secondly, even supposing that the notion of cognitive equilibrium can be consistently defined and is non-empty, the possibility of change in doxastic commitments on mere change in awareness undermines the interest of this notion for the study of doxastic attitudes. For one, if doxastic commitments may change on awareness change, the cognitive equilibrium at which one arrives risks lacking any comprehensible relation with the non-idealised “state of belief and awareness” from which one began. More importantly, if the agent’s cognitive equilibrium may change on changes of awareness, then, contrary to what Levi suggests, invoking equilibria does not allow one to ignore the problem of human cognitive limitations. In constructing an account of the revision of belief or knowledge (understood in terms of cognitive equilibria), one would have to take into account the issue of awareness and awareness change. So, in the absence of stability of doxastic commitments in the face of awareness change, not only does the cognitive equilibrium risk being an abstract object with no recognisable relation to “real” belief states, but it fails to fulfil the role for which it was introduced: justifying ignoring the question of awareness. The equilibrium conception of doxastic states and the changes they undergo requires that such states are unaffected by changes in awareness; it requires IDCA.

The changes of awareness at issue here are changes of awareness *alone*. In natural language, one sometimes speaks of ‘becoming aware’ of something when talking of cases where one just turns one attention to it or gains cognizance of it (as in ‘he became aware of a certain argument against his position’); one also uses the phrase for cases where one focusses on a subject and simultaneously obtains (propositional) knowledge concerning it (as in ‘he become aware that the funds had run out’).<sup>3</sup> The former cases can be thought of as *pure* changes in awareness, whereas the latter cases are *compound* changes: typically, they involve a change of awareness with the addition of some new information (for more on this distinction, see Hill (2010)). Of course, the latter cases pose no threat to the formal models discussed here. Plenty of accounts have been proposed of changes of doxastic states in the face of new information, and they can be applied to such compound changes in awareness. The fact that John’s beliefs (and doxastic commitments) regarding the ozone layer change when he becomes aware that there is a huge

---

<sup>3</sup>Correspondingly, the phrase ‘being aware’ is sometimes used to refer to what the agent has “in mind”, what he pays attention to or is cognizant of (he is aware of the problem and is trying to solve it), but it sometimes also carries an implication of knowledge or belief of that of which the agent is aware (he is aware that he is late). Throughout this paper, the term ‘awareness’ is only used in the former sense; no connotation of knowledge or belief is intended.

hole in it can be accounted for by the aforementioned models by noting that John not only pays attention to a subject of which he was not previously conscious, but he has also received new information concerning it: the models treat the change as resulting from the incorporation of the new information, ignoring the change in attention (the “pure” awareness component). By contrast, the standard models do not and cannot deal with pure changes of awareness, such as the change which occurs when John becomes aware of the issue of whether there is a hole in the ozone layer, without being told whether there is one or not. For good reason: as noted above, these models assume that awareness is of no importance for doxastic commitments.

In order to evaluate the tenability of the equilibrium conception dismissal of the question of awareness, we shall attempt to assess the “cost” of accepting the thesis of independence of doxastic commitments with respect to awareness, which underlies it. Given the flexibility in applications of the traditional models, it is very unlikely that there exists a indubitable proof of the falsity of this thesis: for any apparent counterexample to IDCA, one can probably find some analysis according to which the thesis is not violated. However, such analyses may have some unappealing consequences, and these consequences give a measure of the “direct costs” of holding on to IDCA. Moreover, there may sometimes be good reasons in favour of violating IDCA; since these will have to be rebutted by anyone intent on defending the thesis, they give a measure of the “opportunity costs” of holding onto IDCA. In the following sections, by considering several examples of awareness change which appear to violate IDCA, we identify some direct and opportunity costs of holding onto the thesis.

## **2 Three Examples of Awareness Change**

Consider the following examples.

1. Jim is an air-traffic controller. Ten minutes ago, sitting in front of his screen, he told ground control that flight AA564 was at coordinates X; he believed that AA564 was at X. Since that moment he has not paid any attention to that flight. Now he has to go to the toilet, where he is reminded of the flight by a colleague, who does not know where it is. On being reminded of the flight, his belief has changed: he no longer believes that it is at coordinates X, and vows to check where it is as soon as he gets back to his post.
2. Tom is a doctor. Two years ago, he was the regional specialist in a particular illness: he had read all the literature on the latest cures and medication for that disease. Then he changed hospital and has not seen a patient with this problem since. He has not read one article on the subject, nor given

any serious thought to it, in the intervening period. One day, a patient with the condition arrives in his ward: although Tom remembers that two years prior he believed the drug Z to have good chances of curing the disease, he no longer holds this belief with enough conviction to prescribe the drug, preferring to refer the patient to a doctor who is more up to speed on the disease. His beliefs have changed, though he has not learnt any new information.

3. Rich is a scientist specialising in global warming. In 2008, he made precise predictions about how big the hole in the ozone layer would be at the end of December 2009, though he has not given much attention to this issue since. In January 2010, he is reminded of the issue of the hole in the ozone layer, and asked how big it is. Though he remembers his prediction, he decides to suspend judgement on the issue: after all, there is a research group who will soon have, or already do have, the precise measurements.

These examples have a common structure. At an initial stage, the agent in question has certain beliefs, or doxastic commitments, about a particular issue. It is not unreasonable to say that his state of cognitive equilibrium at that moment contains such commitments (respectively: that AA546 was at X; that drug Z has good chances of curing the disease; that the size of the ozone hole in December 2009 is W square kilometres). In an intervening period, the agent does not consider the issue. There thus seems to be little reason to think that his beliefs, or his doxastic commitments, would change. Finally, at the end of the period, his attention is drawn to the subject: in all cases, just on becoming aware of the issue, the agent changes his beliefs, and what he is doxastically committed to. In all cases, there is apparently nothing irrational about these changes of belief. So these appear to be cases where the agents' doxastic commitments have changed, just because of a change in awareness. *Prima facie*, they are cases where IDCA is violated.

As discussed above, this conclusion is troublesome for traditional models of doxastic states. Were one to want to defend these models, one would have to reject this preliminary analysis of the examples. Two strategies are available. On the one hand, one could argue that beliefs do indeed change when the person in question becomes aware of the issue, but that these are not cases of pure awareness change: there is rather some new information which is learnt and which explains the change in beliefs. On the other hand, one could argue that there is no change in beliefs when the agents become aware of the respective issues, because the beliefs had *already* changed before the moment in question. Let us consider these

strategies.<sup>4</sup>

The first strategy suggests that, in all the cases, the agent learns some new information when he becomes aware of the respective issue (the position of the flight, the drug to use in such circumstances, the size of the hole in the ozone layer). It is difficult to see what this could be: none of the agents appear to receive any new information from the outside world, perception or testimony. One suggestion is that the agent learns that he has not received any information in the intervening period; for example, Tom learns that he has not kept up to speed with the literature in that period. However, this cannot be correct, for suppose that Tom learns at time  $t$  that he will not keep up with the literature in the next two years. Would this mean that, updating on this information, he should relinquish his belief that the drug has good chances of cure already, at time  $t$ ?<sup>5</sup> Another suggestion would be that they learn what time it is at the moment that they become aware of the issue, and that this “new information” triggers the changes in belief. This suggestion is easily dispensed with. It can be assumed that the agents know the

---

<sup>4</sup>We omit explicit discussion of replies which seek to challenge the fact that there has been a change in belief at all, because they are either specific to a particular notion of belief, or they lead to strategies which fall into one of the two categories mentioned in the text. For example, one might doubt that there is any change in full belief in the examples; however, it is difficult to deny that there is a change in the agents’ degrees of belief or confidence in their beliefs, and this is sufficient for our purposes. Indeed, unless indicated, the discussion in this paper holds for all standard notions of belief (so, although at times the language of full belief may be used, this is entirely for expository ease). Alternatively, one might reply to the example by insisting that the “norms for belief” have changed rather than the “beliefs themselves”. However, the main interest of this paper is epistemological, and the epistemologist is interested in whatever sense of the term ‘belief’ there is which satisfies the appropriate norms. This reply does nothing to alleviate the need to explain the change in beliefs in these examples, where ‘belief’ is understood in this sense. Presumably, someone who proposes this sort of reply would wish to defend a modification of standard models of doxastic states which is as radical, if not more radical, than those considered in Sections 3 and 4 below: namely replacing the single notion of belief present in these models with two notions – a “psychological” notion of belief and a notion of doxastic norm – which are considered, and modelled, separately and which combine according to some rule to produce the notion of (“norm-worthy”) belief that is the standard subject of epistemology. Only by such a move can one make sense of the reply mentioned above, via the idea is that the “psychological belief” remains fixed in these examples, whereas the doxastic norm changes. Full discussion of such a move must await a more developed theory along these lines. Let us simply note that, as concerns the change in the “norm-worthy” beliefs, only two options remain if IDCA is to be preserved: either the effect of the norm change on the set of “norm-worthy” beliefs occurs at the moment of awareness change or it has already occurred before the moment when the agent becomes aware of the issue. These correspond to the two strategies discussed in the main text.

<sup>5</sup>This sort of argument can be used to reject any suggestion where what is learnt could be learnt before the moment of the awareness change; such suggestions include proposals about norm changes mentioned in footnote 4. Moreover, taken in tandem with the point made below concerning learning the time, it shows that whatever change is occurring is not a standard case of update; see also Section 3.



time before being made aware of the issue in question, so it does not count as new information. To put the same point more fastidiously, consider modified versions of the examples where the agent in question is told the time just before being made aware of the issue. Either their doxastic commitments change on learning the time, in which case they had already changed before the agent became aware of the issue, and this case is an example of the second strategy for defending IDCA, which will be considered below; or the doxastic commitments in question do not change on learning the time, but rather on being made aware of the issues, in which case the time cannot be the “new information” which explains the belief change. In summary, it is difficult to see how this first prospective defence of IDCA could get off the ground.

That leaves the second strategy, according to which the change in awareness does not induce a change in doxastic commitments because the commitments had *already* changed *before* the moment when the agent became aware of the issue. All the examples involve beliefs which seem to have a natural, in-built “clock”: the idea would be that one’s doxastic commitments change “automatically” as time passes.

This is certainly a coherent way of escaping the conclusion that IDCA is violated; the problem with it is the price to be paid. Such an analysis of the examples sanctions violation of a property shared by all major formal models of belief and belief change, which we call *constant doxastic rest*.

To explain the point, consider example 1. There, the belief concerns a sentence (‘AA546 is at position X’) whose truth value changes in time. One can avoid the suggestion that a change in awareness alone induces a change in belief by explicitly time-indexing the sentences of the relevant language. At the initial time  $t$ , Jim both believed that AA546 is at position X at  $t$  and that AA546 is not at position X at  $t + 10$ . Neither of these beliefs have changed; so the change in awareness has no effect. This is an example of the strategy of proposing that the doxastic commitments (to the sentence ‘AA546 is currently at position X’, in this case) change “automatically” because the object of the doxastic commitment (the sentence) has different content at different moments (due to the indexical ‘currently’) and doxastic commitments to the time-independent contents do not change.

However, this treatment will not work for the other examples. In example 2., it is not the object of the belief (the sentence believed) which is time sensitive but something about the belief itself. The drug in question is as likely to cure the patient now as it was two years ago, so one cannot analyse this example by saying that, two years ago (time  $t$ ), Tom was doxastically committed to ‘drug Z has good chances of curing the patient at time  $t$ ’, but not to ‘drug Z has good chances of curing the patient at time  $t + 2$ ’. Rather, what is needed is that Tom is doxastically committed, at time  $t$ , to drug Z having good chances of curing the

disease, whilst not begin doxastically committed, at  $t + 2$ , to the drug having good chances of cure, *in the absence of any information or “trigger” relevant to this issue in the intervening time*. Hence, in order to defend the position that Tom’s doxastic commitments had already changed before he became aware of the new case, one must give up the following thesis about belief change:

**(Constant doxastic rest)** In the absence of perturbation of an agent’s doxastic commitment with respect to a sentence, his doxastic commitment remains constant.

This principle basically states that belief changes can only occur by perturbations to the doxastic state; the doxastic state of rest, so to speak, is that of constant belief. Of course, the principle as stated is not fully precise, since the notion of ‘perturbation’ of a doxastic commitment remains to be spelled out in more detail. However, it suffices for the purposes of this paper to understand by ‘perturbation’ any deliberate sort of belief change which has been proposed and studied in the literature to date – incorporation of new information considered to be relevant to the sentence in question, moments of inductive inference, suspension of belief on discovery of an anomaly or on the proposal of a competing hypothesis, and so on. (We return to the question of possible perturbations in Section 4.) Understood so, this principle is satisfied, implicitly or explicitly, by all major formal theories of belief and belief change proposed to date. In fact, it is a direct consequence, if not an expression, of the Peircian perspective according to which it is belief change, rather than beliefs themselves, which are to be justified. Since such justification must pass by the receipt of new information or some external signal, or by a deliberate act of (non-deductive) inference of some sort, it is tacitly assumed that, in the absence of these, beliefs remain constant.

If example 2. is analysed as implying a change in doxastic commitment between the initial moment and the moment when Tom is reminded of the disease, then this example violates the principle of constant doxastic rest. For in this example, there is no perturbation concerning the belief about the drug of any sort proposed to date between the two time points, so the doxastic state is at rest; however, the belief has changed, so it is not constant.

Example 3. hammers home the point. In this case, the beliefs at both times concern an issue related to the same moment – the size of the hole in the ozone layer in December 2009 – so one cannot distinguish two time-indexed sentences referring to different moments of time to which Rich has always had different doxastic commitments (as in example 1.). Hence, were one intent on insisting that Rich’s doxastic commitments changed before he was made aware of the issue, one would have to accept that these commitments have changed in the absence of any of the sorts of trigger or deliberate belief change considered to date; one would have to accept that his state of doxastic rest is non constant.

In summary, the thesis of independence of doxastic commitments with respect to awareness can be saved from the threat posed by examples 2. and 3., but at the price of surrendering the principle of constant doxastic rest, which, as noted above, is espoused by all major current models of belief and belief change. But how could an agent's doxastic commitments change in the absence of any perturbing influence? Alternatively, why would doxastic commitments change on merely becoming aware of an issue? In the following sections, we examine the two horns of the dilemma posed by these examples. First of all, we examine the "cost" of retaining IDCA, by considering the consequences of giving up the principle of constant doxastic rest, and in particular the philosophical and technical modifications that would be required in current models of doxastic states to deal with examples such as those given above. Then we consider the "benefits" of abandoning IDCA – or, if you will, the "opportunity costs" of retaining it – by proposing a possible rationale for why changes in doxastic commitments may occur on changes of awareness alone.

### 3 Forgoing Constant Doxastic Rest

In order to retain IDCA, and thus the equilibrium conception of doxastic states, one must abnegate the principle that the only changes in doxastic states occur in the presence of some perturbing influence. In this section, we consider the cost of abandoning this principle, in terms of the modifications required to current models of belief. First of all, we give an argument that violations of constant doxastic rest may be incompatible with models of full belief. It is a surprising, and certainly unwelcome consequence that, if one were intent on holding onto the equilibrium conception, one may be forced to abandon the notion of full belief. Then, in order to gauge what sorts of modifications in current models of belief are required to account for cases where the principle is violated, we consider a natural story which explains how it could be that, in cases such as these, an agent's doxastic commitments change in the absence of any perturbation.

Suppose that you have a model of doxastic states including a notion of full belief, and suppose moreover that you have given up the principle of constant doxastic rest. It nevertheless seems reasonable to retain the following weaker principle:

**(Doxastic continuity)** If an agent is doxastically committed to sentence  $A$  at  $t$ , in the absence of perturbation to this doxastic commitment between  $t$  and  $t + 1$  second, then he is doxastically committed to  $A$  at  $t + 1$  second.

Whereas constant doxastic rest concerns all species of doxastic commitments, doxastic continuity only involves doxastic commitments corresponding to full be-

liefs.<sup>6</sup> And whereas constant doxastic rest concerns any period of time, doxastic continuity concerns only small periods. As such, it seems reasonable, even in the face of rationalisations of non-constant doxastic rest such as those considered below. Even if one admits that beliefs may change in the absence of perturbation, why should Tom's belief that the drug has good chances of cure change from one second to the next (in the absence of perturbation)?

However, by a sorites style argument, doxastic continuity implies constant doxastic rest, for full beliefs. Suppose that an agent is doxastically committed to  $A$  at  $t$ , and there is no perturbation to this belief between  $t$  and  $t + 2$  years. Then the agent is doxastically committed to  $A$  at  $t + 1$  second, by doxastic continuity. Therefore he is doxastically committed to  $A$  at  $t + 2$  seconds, by doxastic continuity. And so on, until  $t + 2$  years  $- 1$  second, whence, by doxastic continuity, the agent is doxastically committed to  $A$  at  $t + 2$  years. So constant doxastic rest holds for the full belief that  $A$ .

If there is a notion of full belief in one's model of doxastic states, doxastic continuity (applied to the full belief) implies constant doxastic rest. The bullet which a defender of the equilibrium model of belief must bite becomes ever less appetising: not only must he relinquish the principle of constant doxastic rest, but in order to do so he must either agree to banish full belief from his model of doxastic states, or he must somehow argue that the full-belief short-term instance of constant doxastic rest – doxastic continuity – is also to be rejected. If violations of constant doxastic rest seem difficult to justify, it becomes almost impossible to imagine how one might defend violations of doxastic continuity. Of course, many will already have doubts about the notion of full belief, but it should nevertheless strike one as strange that they end up being banished because of considerations pertaining to the consistency of the equilibrium conception itself.

Of course, sorites arguments of the sort employed here have received intensive study, especially in recent years, and this is not the place for an in-depth consideration of all the possible replies to an argument of this sort. All the more so that, as the literature on the sorites teaches us, any rejection of the argument will require one to accept some *prima facie* unintuitive position, and the whole purpose of mentioning it here is to emphasise how uncomfortable it may be to reject constant doxastic rest. So let us content ourselves with one remark about a moral to be drawn from an argument of this sort. For those who are willing to give up constant doxastic rest for full beliefs, the argument above emphasises that fact that there must be a moment when the full beliefs change without there being any

---

<sup>6</sup>In the formulation of the principle, we are thus following the convention that, in the absence of a qualification indicating the degree of the belief (or of the doxastic commitment), one is talking about full belief. A version of doxastic continuity can be formulated for other species of doxastic commitments, and in particular for partial beliefs, but, as shall be clear in a moment, it is of at most tangential interest here.

perturbing influence (this is the moment when doxastic continuity is violated). A difficulty for defenders of this position is to explain when the beliefs change, and more importantly why they change when they do; much of the intuition behind doxastic continuity is the sense that such an explanation, if not inexistent, is not forthcoming.

This is an “immediate” cost of taking the defence of the equilibrium conception of belief, and the IDCA thesis, and abandoning the principle of constant doxastic rest. However, anything close to a complete picture of the cost of this move cannot be had without considering what stories could be told about why the principle is to be abandoned. A defender of IDCA cannot simply content himself with the assertion that beliefs may change in the absence of perturbation. And not only because of the immediate counter-intuitiveness of this contention. It calls for a revolution in the modelisation and theorisation of belief and belief change. To gauge exactly how revolutionary it is, to get a glimpse of how much has to be overhauled in current theories, it is necessary to consider what the new sorts of models which do not satisfy constant doxastic rest would look like. To do this, one needs to consider why it could be that doxastic states, left on their own, might “automatically” change. Perhaps the least unnatural story which could be given to explain such changes (we consider another option in the next section) is that the agent *had already committed himself to the change*; the change was automatic, because it was *planned*.

Consider example 2. again, and suppose that we wish to analyse it as involving a change of belief which had already occurred when Tom is reminded of the disease. Then we require that Tom has a doxastic commitment at time  $t$  that the drug Z has good chances of cure, whilst he does not have this commitment at time  $t+2$ , in the absence any new relevant information in the intervening period. Tom’s was a belief with a best before date; perhaps the most natural way of understanding it is as having been conceived as such. That is, by considering that Tom was committed, at time  $t$ , to not holding the belief at  $t+2$  that the drug Z has good chances of cure, in the absence of new relevant information in the intervening period. This would explain why there was a change in beliefs “before” Tom became aware of the issue: he was always committed to the change, so it “happened” before the change in awareness. However, the commitment, at  $t$ , to not holding a belief at  $t+2$ , in the absence of appropriate information in the intervening period, is not a belief in itself, but rather a commitment to believe. This proposal involves something which is alien to current models of belief: commitments at one moment to hold certain beliefs at another moment.

Example 3. can be analysed in a similar way. Rich holds certain doxastic commitments (about the size of the hole in the ozone layer in December 2009) at one moment, which he ceases to hold at another moment. There is no perturbation (new information, discovery of an anomaly etc.) which triggers the retraction of

his belief; rather, he suspends the belief because he knows that there are more reliable methods of deciding the issue than those he used (namely, prediction). Moreover, it was known to Rich when he made his prediction in 2008 that, in the absence of the relevant data or other pertinent information, he could rationally suspend judgement on the issue in January 2010. So, the most natural way of accounting for the change in belief, if one is intent on insisting that it occurred before his change of awareness, is to postulate that he was committed, in 2008, to suspending his belief in 2010, in absence of new relevant information.

Under this analysis, Tom's and Rich's doxastic states involve something of the order of a *doxastic plan* – believe that X holds until moment  $t$ , then suspend belief about X – which they are committed to following unless it is overridden by new information or some other perturbation. However, doxastic plans, and the commitments at one moment to believe at another which they involve, are strangers to current models of doxastic states, because they involve irreducible reference to beliefs at times other than the present. That is, they violate the following widely if implicitly accepted principle:

**(Doxastic actualism)** An agent's doxastic state at time  $t$  consists solely in the beliefs he has at  $t$ .

Taken in tandem with the idea that one is interested not in the beliefs the agent recognises himself as having, but rather in his doxastic commitments, doxastic actualism states that all there is to say about an agent's doxastic state at time  $t$  is contained in what can be said about the doxastic commitments he holds at  $t$ . It is evident that the major models of doxastic states satisfy doxastic actualism: the doxastic state at  $t$  is represented by a set of sentences – those believed at  $t$  – or a probability function over those sentences – giving the degrees of belief of the agent at  $t$  – or more sophisticated versions of similar ideas.

Of course, doxastic actualism does not imply that the contents of beliefs cannot refer to times other than the present. John can believe today that it will rain tomorrow: the fact that the object of the belief refers to another time does not change the fact that this belief is only relevant for his doxastic state today. Nor does it rule out beliefs about beliefs one did or will have at other times: one's current beliefs about what one will believe tomorrow or did believe yesterday count as current beliefs, for it is the object of the belief, not the belief itself, which refers to another moment of time. Moreover, doxastic actualism does not prohibit the agent from having commitments about beliefs at other moments, but it insists that these must follow entirely from his current beliefs. So, for example, present conditional beliefs, taken in tandem with appropriate rules of update, may generate commitments to believe at other moments; if one currently holds a conditional belief in a particular hypothesis given certain evidence, then one has the commitment to

believe the hypothesis at some moment in the future, if one acquires the evidence at that moment. Since such commitments to believe are derived from present (typically conditional) beliefs, they are entirely compatible with doxastic actualism. Doxastic actualism is only violated when there is specific and irreducible reference in the actual doxastic state to one's beliefs or doxastic commitments at times other than the present. This is exactly what happens in cases of commitments to believe or suspend belief at other moments that are not derivable from present beliefs, as is the case in the examples given above. In example 3., it is not that Rich does not believe, in 2008, that the hole in the ozone layer would be as predicted, given that it is January 2010; rather, he plans to suspend the current belief in January 2010. In example 2., Tom does not initially believe that it is not the case that, if he does not receive any information on the subject for two years, then the drug will have good chances of cure, nor does he not believe it to have good chances of cure, given that he does not receive any relevant information for two years; rather, he is committed to suspending his belief about the drug within two years, unless he receives new relevant information in the interim.

In summary, doxastic plans, involving commitments to believe or suspend belief at moments other than the present, are one way to explain how it could be that doxastic commitments change in the absence of perturbation. The key move is to replace actualistic doxastic states – where all there is to a doxastic state at a moment is the beliefs held at that moment – with non-actualistic ones – which may involve commitments to believe or suspend beliefs at other moments. The principle of constant doxastic rest is violated because the standard notion of doxastic state is too restrictive: although the beliefs may change in time without there being any perturbing influence, the *doxastic plans* remain constant in the absence of perturbation.

This is perhaps the best shot one can give at motivating violations of constant doxastic rest. Under it, IDCA and the equilibrium conception can be retained at the price of doxastic actualism; but the costs of dropping this principle are far from slight. A whole new dimension, that of time, needs to be added to belief models; notions of consistent doxastic states will have to be refined accordingly. A new theory of belief formation and belief change is required: it needs to explain, in particular, on what basis new information, which is typically taken to affect current beliefs, can also allow one to judge one doxastic plan to be more appropriate than another which agrees with it on current beliefs. Moreover, abandoning doxastic actualism is not philosophically uncontroversial. Is it necessarily a reasonable normative constraint on rational agents that they not only keep a consistent definite set of beliefs at each moment, but also a consistent definite set of doxastic plans? We conclude that, whilst it is certainly not impossible to defend IDCA from the examples given above, the cost is significant, requiring as it does a major revision of the sorts of models of doxastic states proposed to date.

## 4 Forgoing IDCA

In the previous section, we have considered the price of defending IDCA; this involved, in particular, considering how it could be that one's beliefs could change in the absence of perturbation. Now we consider the other side of the dilemma: that of giving up IDCA. If there are *prima facie* good reasons for thinking that IDCA is violated in a certain situations, then a defender of the thesis will, above and beyond having to wrestle with the consequences detailed in the previous section, have to show what is wrong with these arguments. Hence the following question: why do or would doxastic commitments change on merely becoming aware of an issue?

Consider once again the examples given in Section 2, and suppose that one accepts the analysis according to which IDCA is violated. In example 3., why does Rich change his belief on becoming aware of the question of the size of the hole in the ozone layer? One strong intuition is that he is *re-evaluating* his old belief concerning the size of the hole, and on so doing, finds that the belief now has insufficient warrant. Re-evaluation of existing beliefs (in particular beliefs formed at previous moments) is, as we shall see, one possible explanation for why doxastic commitments may change on mere changes in awareness.

First of all, it is crucial to note that the act of belief re-evaluation is a *bona fide* species of change in doxastic commitments. In the examples, there is nothing contradictory in the agents retaining their existing beliefs and not carrying out the specified belief changes. Although in many cases, on re-evaluating one's beliefs, the trigger and the motor for the re-evaluation may be deductive – identification and resolution of an inconsistency in one's beliefs – this is *not* always the case; it is not the case in the examples given above. Hence, even if one admits a consistent and non-empty notion of cognitive equilibrium, re-evaluation of beliefs is not something which is necessarily already “taken into account” in the shift from the “real” state of belief and awareness to the cognitive equilibrium. In 2008, whilst aware of the issue, Rich held the doxastic commitment that the ozone hole was of size W at the end of 2009; in 2010, he was aware of the issue but no longer held this doxastic commitment: this is a change in his cognitive equilibrium, not a shift from a non-equilibrium state to an equilibrium one. Therefore, even for the epistemologist who wishes to trade only in cognitive equilibria and to ignore the “imperfections” of “real” belief states, re-evaluation is a *bona fide* species of belief change. Needless to say, re-evaluation may be a rational way of changing beliefs, in appropriate situations, and so is of potential interest to an epistemologist.

Furthermore, re-evaluation is a species of change which has been largely if not completely ignored in the literature. The sorts of belief change present in the examples above are contractions (suspensions of belief), but they do not seem to fall into any of the categories of contractions identified to date. As already empha-



sised, they are not contractions in the face of inconsistency (what Levi (1991) calls ‘coerced contractions’), but they do not fall under the only other sorts of contraction which have been identified, namely contractions in the face of anomalies and in order to accommodate newly proposed theories or hypotheses (the only sorts of ‘uncoerced contraction’ identified by Levi (1991, §4.9)). For these latter sorts of contraction always have some sort of “external informational trigger”, be it the discovery of an anomaly or the introduction of a new theory or hypothesis. No such thing is present in the current examples; only the awareness of the issue has changed. In fact, this difference may point to the reason why re-evaluations have been widely disregarded to date. In the Peircian perspective, which is highly influential if not dominant, the problem is to justify changes in belief. Standardly, this justification passes by some sort of trigger, be it incoming new information, introduction of a new theory, discovery of an anomaly or contradiction; these correspond to the perturbations that we met in Section 2. It is natural that re-evaluation be ignored by those working in this perspective if no answer is forthcoming to the question: what is the trigger for re-evaluation of beliefs? However, this question need not go unanswered.

Note first of all that, typically, the identification of a trigger for a belief change not only provides a *justification* for the change, but it also fixes the *timing* of the change. On learning new information, the agent is not only justified in altering his beliefs in the light of the information, but he is also justified in altering those beliefs *at the moment* when the new information arrives, or as soon after it as possible. That the question of the trigger for a belief change and that of the timing of the change are related is, of course, completely natural: it is a reasonable principle that one should undertake a belief change at the first available opportunity and in all the standard cases cited above, this is upon the arrival of the trigger for the change.

In the light of this relationship between the trigger for a change and the timing of the change, a trigger can be proposed for re-evaluations of belief. For there are often firm intuitions concerning the timing of re-evaluations: in certain situations at least, one should re-evaluate one’s belief at the first available opportunity. But, more often than not, the first available opportunity is when one becomes aware of the issue with which the belief is concerned. In the examples above, Tom and Rich re-evaluate their beliefs on becoming aware of the issue of the chances of cure and the hole in the ozone layer; if they had become aware of the respective issues ten minutes earlier, they would have re-evaluated their beliefs ten minutes earlier. Given the relationship between the timing of a change and its trigger, we can conclude that changes in awareness are potential triggers for re-evaluation.

The idea that a change in awareness may be a trigger for the act of re-evaluation of one’s beliefs sheds light on the question of why the thesis of independence of doxastic commitments with respect to awareness may be violated: one case of

violation is when the change of awareness triggers a re-evaluation which leads to a change in belief. It explains moreover why re-evaluation of beliefs have been ignored for so long: one major trigger for these sorts of changes has been “disqualified” from consideration by the equilibrium conception argument for ignoring awareness. We conclude that there is good reason for abandoning the thesis of independence of doxastic commitments with respect to awareness, on the basis of the recognition of a species of belief change – re-evaluation – which may be triggered by changes of awareness. And the principle of constant doxastic rest is preserved, once this as yet unconsidered sort of perturbation is taken into account.

Let us stress that it is not being suggested that changes in awareness are all there is to say about when one should re-evaluate; it suffices for our purposes that changes of awareness play an integral part in some triggering of re-evaluations of beliefs. A full discussion of the appropriate circumstances for belief re-evaluation, not to mention of how beliefs are to be re-evaluated, is beyond the scope of this paper. Our only contention is that, though a full specification of the circumstances will doubtless be more intricate than the simple remarks made above, change in awareness will play a role. So, for example, one suggestion is that certain categories of beliefs should be re-evaluated at the first available opportunity after a sufficiently long period of not being explicitly used or having received explicit attention; of course, this boils down to re-evaluating the beliefs in these categories on becoming aware of them after an appropriate period of unawareness.

Introduction of the act of re-evaluation of beliefs thus allows one to sanction changes in doxastic commitments which occur at moments of awareness change, and retain the principle that in the absence of perturbation (which may include changes in awareness), beliefs remain constant. It gives a principled reason, acceptable even to an epistemologist concerned only with doxastic commitments and not with the “imperfections” of “real” belief states, for thinking that IDCA may not hold in general. Moreover, it gives an idea of what a theory that does not espouse IDCA would look like. Above and beyond involving awareness in the model of doxastic states and in considerations of change of doxastic states, it would include an act of re-evaluation of beliefs, which could be triggered by changes of awareness, under suitable conditions.<sup>7</sup> Some important steps in the direction of such a theory have already been taken, with the development of models of awareness and of changes involving awareness (see Fagin and Halpern (1988); Heifetz et al. (2006); Hill (2008, 2010) for some examples); other elements, such as the conditions for re-evaluation, remain to be explored. In the light of this, the project of developing a full theory of belief and belief change which does not rely on IDCA and which is of epistemological interest, does not seem, neither

---

<sup>7</sup>Formally speaking, modelling such an act need not require more than the standard tools already developed by theories of belief revision and awareness change.

philosophically nor technically, as daunting as might have been first imagined. Certainly, it is not clear that this road out of the dilemma posed by the examples above is more difficult than the alternative considered in the previous section.

Before closing, let us note that the introduction of the operation of re-evaluation of beliefs does not lead inevitably to the rejection of IDCA and the preservation of constant doxastic rest. This depends rather on the answer one gives to the question: when should one re-evaluate beliefs? The answer given above – on the first available opportunity under certain conditions, and thus in particular on certain instances when one becomes aware of the belief – leads to the violation of IDCA, but the preservation of constant doxastic rest. However, another answer can be given to the question, with opposite consequences: beliefs should be re-evaluated *at every moment*.

This answer provides a different analysis of the examples in Section 2. In Example 2., it is not that Tom re-evaluates his belief about the chances of cure on becoming aware of the issue; rather, he has already changed his belief, following a re-evaluation at some previous moment. By maintaining that beliefs are continually re-evaluated, one can avoid the conclusion that IDCA is violated in these examples, because the beliefs in question had *already* changed before the awareness change. However, these previous re-evaluations must have occurred at moments when the agent was unaware of the belief which was being re-evaluated. So this position amounts to the *automatic* re-evaluation of beliefs, at every moment, including moments of unawareness of the beliefs being re-evaluated. As such, it violates the spirit, if not the letter, of the principle of constant doxastic rest. This principle supposes a meaningful state of doxastic rest, where there is no perturbation, and demands that in such cases beliefs remain constant. The thesis of continual re-evaluation either violates the letter of this principle – if the omnipresent re-evaluations are not understood to be perturbations, then beliefs can change in a vacuum, in the process of continual re-evaluation – or it violates the spirit of the principle – if the re-evaluations are considered as perturbations, then by the thesis of continual re-evaluation, doxastic states are never left unperturbed (not even in principle). By maintaining that all beliefs are or should be re-evaluated at every moment, one escapes the conclusion that IDCA is to be rejected, and relinquishes instead the principle of constant doxastic rest.

The proposal of continual re-evaluation suffers from several of the disadvantages identified in the previous section with positions which abandon constant doxastic rest. In endorsing such a position, one will have to decide between rejecting the principle of doxastic continuity or abandoning the notion of full belief. Formal models which incorporate the position would resemble in many ways the models which abandon doxastic actualism, and would require a similar overhaul. In both cases, some theory would be required of how and when beliefs change “automatically” in the absence of standard perturbing influences: in the previous

case, these changes are understood to be the result of prior planning, whereas under the current proposal, they would be carried out automatically.<sup>8</sup>

But it is philosophically that the proposal of continual re-evaluation is most controversial. Why *should* an “ideal”, fully rational agent re-evaluate all of his beliefs at every moment? Isn’t this too much to ask for, even of him? Such a philosophical position can doubtless be defended; in doing so, one effectively abandons the Peircian perspective according to which it is changes of belief, rather than beliefs themselves, which are the main objects of attention and of justification. Re-evaluating one’s beliefs at every moment is tantamount to “re-justifying” them continually, and this flies in the face of the Peircian tradition. Indeed, under the thesis of continual re-evaluation, the classical problem of belief change is of minimal importance, becoming a simple corollary of the theory of re-evaluation (at appropriate moments, on re-evaluation, one may find that new information has arrived since the re-evaluation a second before). The argument in this paper has been to point out some tensions between the sidelining of the notion of awareness and the understanding of changes of belief. Of course, this argument loses its bite if the question of belief change is effectively relegated; this is basically what happens if one maintains the thesis of continual re-evaluation. In this case, the issue becomes much deeper, involving as it does basic questions concerning the scope and goal of epistemology. It thus suffices to note that for any epistemologist who desires to retain some aspects of the Peircian perspective and the importance of belief change, the continual re-evaluation option is a very difficult road indeed. In terms of cost, it does not fair better than the rejection of doxastic actualism considered in Section 3.

## 5 Conclusion

The models of doxastic states which are widely used today in formal epistemology to analyse attitude change and decision are relatively simple and attractive. The purpose of this paper is to highlight the price paid for this simplicity.

In particular, it is commonly assumed that, since these models concern equilibrium states of belief, the question of awareness can be ignored. The examples above give reason to reflect upon the wisdom of this methodological choice. They involve cases where pure changes of awareness appear to lead, in a not unreasonable way, to changes in doxastic commitments. The only way of escaping from this conclusion leads one to give up another fundamental assumption of the

---

<sup>8</sup>Indeed, to the extent that the reasons for revising a belief on re-evaluation can be foreseen, the change can be planned at the outset; hence there may be significant similarities between the theories developed under the two proposals.

models: namely, that in the absence of perturbation, one's doxastic state remains constant.

The examples leave defenders, and users, of these models with a dilemma. If they wish to retain the equilibrium conception of belief, and the associated argument for ignoring awareness, then they must abnegate the principle that at rest, doxastic states remain constant. Beyond the inconvenience of rejecting an intuitively attractive principle, this option implies considerable philosophical and technical sacrifice: the notion of full belief may have to be abandoned, and, under the most natural explanation of why doxastic states are not constant at rest, a fundamental principle underlying all major models of beliefs – that all that is relevant to one's current doxastic state are one's current doxastic commitments – must be abandoned.

On the other hand, if they wish to avoid these sacrifices, they must accept that awareness changes may induce changes in doxastic commitments, and bring awareness into the modelling of doxastic states, with appropriate refinement, if not rejection, of the equilibrium conception. However, once one recognises the *sui generis* act of re-evaluation of beliefs, there may be epistemologically valid reasons for thinking that doxastic commitments may sometimes change on mere changes in awareness; consequently, an epistemologically relevant theory of belief in which awareness plays a role can be envisaged.

In conclusion, ignoring awareness is not as straightforward as it at first seemed. Not only may there be good epistemological reasons for taking it into account, but the dismissal of awareness does not come for free: it in itself entails other modifications of standard models of belief.

## References

- Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 4:1236–1239.
- Bradley, R. (2007). The kinematics of belief and desire. *Synthese*, 56:513–535.
- Dekel, E., Lipman, B. L., and Rustichini, A. (1998). Standard state-space models preclude unawareness. *Econometrica*, 66:159–173.
- Fagin, R. and Halpern, J. Y. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34:39–76.
- Gärdenfors, P. (1988). *Knowledge in Flux : Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA.

- Halpern, J. Y. and Rêgo, L. C. (2007). Generalized Solution Concepts in Games with Possibly Unaware Players. In *Proceedings of the Eleventh Conference on Theoretical Aspects of Rationality and Knowledge*, pages 253–262.
- Heifetz, A., Meier, M., and Schipper, B. (2006). Interactive unawareness. *J. Econ. Theory*, 130:78–94.
- Hill, B. (2008). Towards a “Sophisticated” Model of Belief Dynamics. Part II: Belief Revision. *Studia Logica*, 89(3):291–323.
- Hill, B. (2010). Awareness dynamics. *Journal of Philosophical Logic*, 39:113–137.
- Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. College Publications, London. New Edition, 2005.
- Jeffrey, R. C. (1972). *The Logic of Decision*. University of Chicago Press, Chicago, 2nd edition.
- Levi, I. (1980). *The Enterprise of Knowledge. An Essay on Knowledge, Credal Probability and Chance*. MIT Press, Cambridge, MA.
- Levi, I. (1991). *The Fixation of Belief and its Undoing. Changing Beliefs through Inquiry*. CUP, Cambridge.
- Ramsey, F. P. (1931). Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*. Harcourt, Brace and Co., New York.
- Savage, L. J. (1954). *The Foundations of Statistics*. Dover, New York. 2nd edn 1971.
- Stalnaker, R. C. (1984). *Inquiry*. MIT Press, Cambridge, MA.