11

Dynamic Beliefs and the Passage of Time

Darren Bradley

1   Introduction

How should our beliefs change over time? Much has been written about
how our beliefs should change in the light of new *evidence*. But that is not
the question I'm asking. Sometimes our beliefs change without new evi-
dence. I previously believed it was Sunday. I now believe it's Monday. In
this paper I discuss the implications of such beliefs for philosophy of mind.
I will argue that two-dimensionalism (e.g. Perry 1979) about the objects of
belief are supported over one-dimensionalism (e.g. Lewis 1979) for two
related reasons. First, two-dimensionalism gives us a more natural account
of belief retention. Second, the extra complexity of two-dimensionalism
turns out to be independently motivated by confirmation theory. So if the
argument is correct, it will be an instance where our epistemology informs
our philosophy of mind.

2   The Propositional Theory of Belief

The propositional theory of belief states that when an agent believes some-
thing, he is standing in a certain relation to a proposition; namely, the rela-
tion of believing it. This theory has two features that are in tension.[1]

First, the objects of belief, propositions, are eternally true or false. That
is, if true at a time, they are always true, and if false at a time, they are al-

ways false. They do not vary in truth-value like the sentence 'It is Tuesday', which is true one day, false another.

Second, if a rational agent believes proposition P, but doesn't believe proposition P', then P and P' are different propositions. This feature individuates propositions in terms of their *cognitive significance*, which can be thought of as the role the belief plays in the life of the agent.

The tension between these features is brought out by self-locating beliefs. John Perry (1979) tells the story of how he followed a trail of sugar around the supermarket looking for the person who was making a mess. After walking in a circle he realized that he was the person making a mess and bent down to fix the bag of sugar. But what was this belief that he discovered? It can be expressed as 'I am making a mess'. But this straightforward belief presents a problem for the propositional theory of belief.

The sentence 'I am making a mess' doesn't have the first feature mentioned above – it is not eternally true or false. Instead, it is true for one person and false for another. So the belief Perry discovered is not completely expressed by this sentence.

Advocates of the propositional theory of belief might respond that 'I' is short for some concept which always picks out John Perry. Suppose this concept is expressed by *John Perry*. The proposition learnt would then be completely expressed by '*John Perry* is making a mess'. But here we run into of the second feature of belief – that if a rational agent agrees with proposition P, but doesn't agree with proposition P', then P and P' are different propositions. Imagine John Perry had amnesia and didn't remember who he was. Then he might believe the proposition expressed by '*John Perry* is making a mess' while not believing the proposition expressed by 'I am making a mess'. So rather than expressing the proposition believed, adding a concept that always picks out John Perry would turn it into a different one.

Perry argues convincingly that there is no way to turn the belief expressed by 'I am making a mess' into something that fits the propositional theory of belief. Although Perry focusses on indexicals, they are not needed for the problem to be raised. Take Salmon's (1989) example 'Frege is writing'. This sentence changes in truth value as time passes, so the sentence alone cannot completely express a proposition. And adding a concept that picks out a specific time – Frege is writing at t – changes it into a different proposition, for someone who has lost track of the time may accept 'Frege is writing' but not accept 'Frege is writing at t'.

So there is no point trying to fix the problem by trying to get rid of the indexicals. The problem is due to a fundamental tension between taking the objects of belief to have eternal truth-values and taking them to be individuated by cognitive significance. The cognitive significance of a proposition varies according to where the agent is located, but the truth-value does not. Most philosophers have given up one of the features mentioned above.

Temporalists hold that some beliefs have a truth-value only relative to a time. So for example 'it is Tuesday' could completely express a belief – the belief being true relative to Tuesday. So temporalists keep the second feature, that beliefs are individuated by cognitive significance, but give up the first, that they have eternal truth-values. Eternalists hold that all beliefs have eternal truth-values. So eternalists keep the first feature (eternal truth-values) but give up the second (individuation by cognitive significance). Two-dimensionalists[2] posit both types of objects of belief.

It is widely agreed that we need a notion of beliefs individuated by cognitive significance, so in practice, and in this paper, eternalism gives way to two-dimensionalism. So the debate is between two-dimensionalists who posit both temporal and eternal beliefs, and one-dimensionalists who posit only temporal beliefs. This paper argues for two-dimensionalism and against one-dimensionalism.

I think we need the eternal beliefs that two-dimensionalism supplies for two related reasons. First, temporalism does not allow an ontology of *dynamic beliefs* that are retained as time passes. Second, temporalism obscures an important distinction between two rules of belief update.

For concreteness, my main target will be the most influential temporalist theory (Lewis 1979). As a stalking horse, I will use Perry's (1979) well-known two-dimensionalist theory. Let's first lay out Lewis and Perry's theories, both of which were motivated by the self-locating beliefs mentioned above.

3 Lewis's Temporalism

Imagine a picture of all the possible worlds, spread out across logical space. Eternalists can think of a belief as locating oneself in a set of these possible worlds. When you believe grass is green, you believe that you have the property of being in a possible world where grass is green. You are locating yourself in logical space. For eternal beliefs the boundaries of where you are locating yourself match the boundaries of the possible worlds. But why should we restrict ourselves to such beliefs? Lewis argues there is no reason.

> We can have beliefs where we can locate ourselves in logical space. Why not also beliefs where we locate ourselves in ordinary time and space? We can self-ascribe properties that correspond to propositions. Why not also properties of the sort that don't correspond to propositions?…Why not?

---

2 See Kaplan (1989), Stalnaker (1978) for two-dimensional theories of meaning. The framework has been applied to beliefs most vigorously by Chalmers (2002 and many other places). Their arguments are usually based on issues regarding assertion and modality. My arguments are based on mereology (first) and epistemology (second).

No reason! We can and do have beliefs where we locate ourselves in ordinary time and space. (1979 p. 519)

The problematic cases above were just such beliefs. To allow such beliefs Lewis suggests that 'to believe…is to self-ascribe the corresponding property' (*ibid.* p.518). So to believe that I am making a mess is to self-ascribe the property of making a mess. And to believe that Frege is writing is to self-ascribe the property of being in a world, at a time, when Frege is writing. Call these *attitudes de se*.

Lewis denies that beliefs have an eternal truth value, and holds that they are individuated by cognitive significance. Thus, beliefs will figure in common sense psychology, as they will help explain and predict behaviour. What's important for us is what this leaves out. There is no place in Lewis's theory for a category of eternal beliefs. I will argue that this is a mistake.

4   Perry's Eternalism

Perry (1979) introduces a two-dimensional account. Beliefs have a content and a role.[3]

The content is the eternal belief. When I say 'John Perry is making a mess' and John Perry says 'I am making a mess', the content of the beliefs expressed is

< John Perry, making a mess, t >.

The content has the first feature of propositions – it is eternally true (or false).

But the content misses out a key feature of belief – its role in common sense psychology. One cannot generally tell merely from the content, what should be done about it; we also need to know *the way in which it is believed.* So a second dimension of belief is posited. Perry believes the above content with the *role* expressed by 'I am making a mess'. I believe it with the role expressed by 'You are making a mess'. Perry's belief causes him to bend down and fix the bag of sugar. Mine causes me to tell him he's making a mess. The role for Perry is analogous to all beliefs for Lewis – they figure in common sense psychology and are individuated by cognitive significance. But on Perry's account, each belief consists of a content as well as a role.

Which theory is better? This depends on whether the extra complexity of Perry's theory buys anything. Lewis has a unified account. All objects of belief are self-ascribed properties. Perry has a two-dimensional account. He can do everything Lewis can do, but he has a more complicated way of do-

---

[3] Kaplan (1989), influenced by Perry, also has a well known notion of content. But the issue is complicated because Kaplan is a temporalist, so I focus on Perry. Like him, my focus is on belief rather than assertion.

ing it.[4] I will argue that Perry's theory buys us unified beliefs. For Lewis, the *theory* of belief may be unified, but the *beliefs* are not. For Perry, the *theory* is less unified, but the *beliefs* are more unified. Perry allows us an ontology of *dynamic beliefs* – beliefs which persist through time and are apprehended with different roles. I will then argue that the extra complexity of Perry's theory is independently motivated, and is therefore a virtue.

5   Evans' Argument for Dynamic Beliefs

Let's start with Evans' (1990) arguments for dynamic beliefs, which I find inconclusive, but which will help set up my arguments. Evans argues that the basic unit of belief must be something that is retained over time.

> 'The thought units [beliefs] of the atomist [temporalist] are not coherent, independent thoughts at all, but, so to speak, cross-sections of a persisting belief state which exploits our ability to keep track of a moment as it recedes in time.' (1990 p.86)

If this is right, this would refute not just temporalism, but two-dimensionalism, as two-dimensionalists hold that temporal beliefs are coherent independent thoughts. Why does Evans hold this? His reason is given in the previous sentence, so let's extend the quote. Evans claims that:

> '[1] a capacity to keep track of the passage of time is not an optional addition to, but a precondition of, temporal thought. [2] If this is so, the thought units of the atomist are not coherent, independent thoughts at all, but, so to speak, cross-sections of a persisting belief state which exploits our ability to keep track of a moment as it recedes in time.' (ibid.)

In a moment I will deny that [2] really follows from [1]. But let's first consider [1] - why should we think that a capacity to keep track of time is a precondition of temporal thought?

I can find two arguments in Evans (plus a third passing suggestion we'll come back to). Here is the first.

> 'No one can be ascribed at t a belief with the content 'It is now A', for example, who does not have the propensity as time goes on to form beliefs with the content 'It was A just a moment ago, 'it was A earlier this morning, 'it was A yesterday morning'' (1990 p.86)

I don't think this is true. For a real-life counter-example, Clive Wearing has a memory of less than 5 minutes, due to a virus that damaged his brain in 1985. For a few minutes at a time, he is perfectly normal, except for his lack of memories. If you tell him it is raining outside, he will believe you, and repeat it back if asked what the weather's like. But he has no capacity

---

[4] Lewis: 'Whenever I say someone self-ascribes a property X, let Perry say that the first object of his belief is the pair of himself and the property X. Let Perry say also that the second object is the function that assigns to any subject Y the pair of X and Y' (p.537).

later on to form the belief that it was raining this morning, as by then, he will have forgotten it. Presumably Evans has to say that Wearing does not really believe that it is now raining. This seems implausible.[5]

Evans offers a second argument that a capacity to keep track of the passage of time is a precondition of temporal thought. This is based on an analogy with space. To show that our beliefs are based on our ability to keep track of time, he argues that our beliefs are based on our ability to keep track of space. He gives the example of objects moving, but not so fast that we can't keep track if we watch them. Suppose we start with a belief that one of the objects is valuable. On Perry's conception (that Evans is defending), the belief that the object is valuable persists over time. On the atomistic conception, we have a sequence of different beliefs, and

> 'it ought to be possible to have just one of the members of the sequence no matter which others accompanied it i.e. in the absence of any capacity to keep track of the object. But if that ability is missing, it is not possible for a subject to have a thought about an object in this kind of situation at all.' (1990 p.87)

Let's grant that Evans is right about this case. We won't know which object is valuable unless we remember which object was valuable a moment ago. But it's not clear this proves the point. While we sometimes need to track objects carefully, sometimes we don't, in which case Evans's argument fails to generalize. If the valuable object were the only shiny one, it wouldn't matter if we had failed to keep track of the object.[6] We could still have any of the atomic beliefs expressible at some time as '*that* (shiny) object is valuable'. (This would be analogous to the temporal belief that it is now raining.) Such cases seem to lend support to the idea that we should have an atomic conception of belief just as Evans's example lends support to the dynamic conception.

So Evans' arguments for his view that a capacity to keep track of the passage of time is a precondition of temporal thought [1] are inconclusive. But even if true, [1] does not support the conclusion that 'the thought units of the atomist are not coherent, independent thoughts' [2].[7] Let's grant [1] that a capacity to keep track of the passage of time *is* a precondition of temporal thought. So we grant that

> No one can be ascribed at t a belief with the content 'It is now A', for example, who does not have the propensity as time goes on to form beliefs with the content 'It was A just a moment ago, 'it was A earlier this morning, 'it was A yesterday morning'' (1990 p.86)

---

[5] Evans could perhaps adopt a functionalist or dispositionalist account of belief (Schwitzgebel 2002) and argue that the propensity to form future beliefs is one of the essential dispositions of temporal beliefs. But see fn 8.

[6] Compare Branquinho 1999.

[7] Thanks to Wolfgang Schwarz here.

Nevertheless, 'it is now A' and 'it was A just a moment ago' could be coherent independent thoughts. To see this, suppose that due to idiosyncratic features of human psychology, no-one could be ascribed a belief with the content 'I fear X' who does not have the propensity as time goes on to form beliefs with the content 'I hate X'. This doesn't imply that these are not coherent independent thoughts. So it does not imply that they are just two sub-sections of the same thought. Indeed functionalism and dispositionalism[8] entail that there are such constitutive connections between beliefs, but this does not entail that the beliefs are cross-sections of a single holistic belief state.

Nevertheless, Evans does make an argument in passing that I find much more promising.

> 'One belief cannot give rise to another by any *inference*, since the…belief[9] that would be required to underwrite the inference is not a thinkable one; no sooner does one arrive in a position to grasp the one side of the [belief] than one has lost the capacity to grasp the other'. p.86

I think there are a couple of responses the temporalist could make here. First, they need not say that the later belief is justified by the earlier one. Instead, they could say that both beliefs are justified by some earlier experience, and the persisting memory of that experience. Or perhaps they could say that the later belief is justified by the memory of the earlier belief. But this argument of Evans does draw attention to the fact that rather than an ontology of beliefs that persist through time, the temporalist posits a multitude of beliefs that only last for a limited period of time, before being replaced by new beliefs. This is the major weakness of temporalism, or so I will argue.

6 Dynamic Beliefs

Lewis asked what happens when we replace propositional attitudes by attitudes *de se*.

---

[8] Thus suggesting the functionalist or dispositionalist approach turns out not to have been such a friendly suggestion after all.

[9] I have omitted two references to 'identity beliefs' in this passage. Presumably Evans meant beliefs of the form 'A is identical to B' where A and B are atomic beliefs. But the atomic belief theorist doesn't think there is an *identity* between A and B. In fact he is committed to denying the identity of A and B, which is the source of the disagreement with the dynamic belief theorist. I think Evans should have used some justifying relation that falls short of identity. Thanks to Elliott Sober for spotting this.

'Answer: Very little. We replace the space of worlds by the space of cen-
tered worlds….All else is as before.'[10] (1979 p.534)

But things are not so simple. *De se* beliefs can change in a way that
eternal propositions cannot. The reason is that *de se* facts change in truth-
value as time passes, so *de se* beliefs must be adjusted to keep up. The ques-
tion is whether this should be modelled as a case of belief change. Consider
an agent who sincerely utters 'today is Monday' (on Monday) and 'yester-
day was Monday' (on Tuesday). Do these sentences express the same be-
lief? Lewis says no. Having the former belief is self-ascribing the property
of being temporally located on a Monday, while having the latter is self-
ascribing the property of being temporally located on a Tuesday. They are
different properties, so they are different beliefs. But surely this is not really
a case of belief change. This consideration is the crux of my argument so
let's see what we can say to support it.

Consider that 'changing one's beliefs' is more naturally expressed in
English as 'changing one's mind'. And surely no-one changes their mind
when they change from believing it is Monday to believing it is Tuesday.
Lewis's theory gives us the implausible result that we change our minds as
the clock reaches midnight. So Lewis's theory is too fine-grained. Instead,
we need a theory on which the belief is retained – we need an ontology of
dynamic beliefs – and such an ontology is supplied by Perry's content.[11]
(Of course, nothing stops us from introducing 'changing one's beliefs' as a
technical term and stipulating that our 'beliefs change' as the days pass, but
this would not be appealing to someone like Lewis who's trying to vindi-
cate common-sense psychology.)

And as mentioned earlier the problem is not restricted to explicitly tem-
poral beliefs such as 'it is Monday'. Many of our beliefs locate us in time
implicitly. We cannot believe 'Frege is writing' without making implicit
reference to our temporal position, so for Lewis we cannot retain this belief
over time. Consider what Lewis's theory says about the belief(s) expressed
by 'Frege is writing' at one time and by 'Frege was writing' at a later time.
Has the same belief been expressed? Lewis's theory says no. We are ascrib-
ing different properties in the two cases. We are first self-ascribing the

---

[10] To be fair to Lewis, he makes this much quoted comment in the context of discussing *de-
cision* theory. He makes no explicit comment about *confirmation* theory. See
http://www.umsu.de/wo/2010/563 for more discussion.

[11] Richard (1981) offers a similar argument for eternalism. But notice we need not just eter-
nalism, but an eternalism that allows for dynamic beliefs (see fn. 12). Kaplan's (1989) discus-
sion of cognitive dynamics presupposes that beliefs are usually retained: 'Suppose that yester-
day you said, and believed it, "It is a nice day today." What does it mean to say, today, that you
have retained that belief?...Is there some obvious standard adjustment to make to the character,
for example, replacing *today* with *yesterday*? If so, then a person like Rip van Winkle, who
loses track of time, can't retain any such beliefs. This seems strange.' p.537-538.

property of being in a world at a time when Frege is writing. We are then self-ascribing the property of being in a world at a time before which Frege was writing. They are different properties, so they are different beliefs.

The depth of the problem can be brought out by considering the trouble Frege ran into with indexicals. Frege resorted to the view that all beliefs involving 'I' were incommunicable.

> 'Dr. Gustav Lauben says 'I have been wounded'. Leo Peter hears this and remarks some days later, 'Dr. Gustav Lauben has been wounded'. Does this sentence express the same thought [belief] as the one Dr. Lauben has uttered himself?' (1967 p.24)

Frege concludes that it does not. The reason is that Frege (1892) wanted to individuate beliefs by cognitive significance. Someone could accept 'I have been wounded' without accepting 'Dr. Lauben has been wounded', so these sentences have different cognitive significance, so they must express different beliefs. But Frege also wanted beliefs to have eternal truth-values, so he denied that 'I have been wounded' expresses a single belief when uttered by different people (as the sentence would sometimes be true and sometimes false). With these constraints, communicating the belief expressed by 'I have been wounded' is problematic. It has not been communicated if someone else accepts 'Dr. Lauben has been wounded', nor if someone else accepts 'I have been wounded'. So Frege resorted to incommunicable senses – no-one else could grasp the belief Dr. Lauben expressed with the words 'I am wounded'. This in itself is an unhappy conclusion. But to make things even worse, suppose that, *per impossibile*, someone changed identities. Then they could no longer believe anything they had previously believed using 'I'; the later 'I' would refer to a different person, so the belief would be different. Of course this cannot happen with persons, but it can happen with times.

The analogous view regarding 'now' is that someone cannot express a belief on one day using 'now' and then express the same belief at a later time. The same applies for other terms such as 'today' and 'yesterday'. And implicitly temporal beliefs such as expressed by 'Frege is writing' also cannot be expressed at a later time.[12]

But when it came to time, even Frege could not accept such a view

> If someone wants to say the same today as he expressed yesterday using the word 'today', he must replace this word with 'yesterday'. Although the

---

[12] Kripke (2008) interprets Frege as saying exactly this; sense is both eternal and atomistic i.e. cannot always be grasped at a later time. Kripke implies that the famous quote of Frege just below in the main text is confused; '[i]f Frege really means that we have expressed literally the same thought again, it is very hard for me to see how to reconcile this assertion with his other doctrines.' p. 204. Kripke argues that the resulting incommunicable senses are independently motivated. My arguments for belief retention also cut against this unusual variant of eternalism. In contrast, Evans (1990) argues that the same sense *is* expressed on different days p. 208.

thought [belief] is the same its verbal expression must be different. (1967 p.24)

Frege thus seems to reject his early view (1892) that beliefs are individuated by cognitive significance in favour of a view that allows belief retention, and I suggest he would have been right to do so.

Let's press the point a little further. The view which I am attributing to early Frege and Lewis, that we countenance only sense (cognitive significance) and reference in our theory of belief, leaves out dynamic beliefs. To see how bad this is, consider the beliefs expressed by:

1. Snow is white

2. London is in England

These have the same reference (true) and different senses. Compare the beliefs expressed by:

3. Today is Monday (said on the 6th)

and

4. Yesterday was Monday (said on the 7th)

They also have the same reference (true) and different senses. Which means that for (early) Frege and Lewis, 1 and 2 are no more similar than 3 and 4. But this has clearly left something out. It has left out that 3 and 4 express the same belief – a dynamic belief. This completes my main argument against Lewis's view. Let's now consider an objection.

My argument relies on intuitions about belief identity. But an objection to the importance of such intuitions can be extrapolated from Lewis's work. Lewis (1980b) argues that Kaplan's (1989) notion of same-saying is insignificant. We can find an analogous argument against our notion of same-believing. I have changed 'said' to 'believed' in the following quote:

(1) I believe 'I am hungry'. You simultaneously believe 'You are hungry'. What is believed is the same

(2) I believe 'I am hungry'. You believe 'I am hungry'. What is believed is not the same….

(3) I believe on 6 June 1977 'Today is Monday'. You believe on 7 June 1977 'Yesterday was Monday'. What is believed is the same….

I put it to you that not one of these examples carries conviction. In every case, the proper naïve response is that in some sense what is believed is the same for both sentence-context pairs, whereas in another – equally legitimate sense – what is believed is not the same. (adapted from Lewis 1980b / 1998 p.41)

Let's grant this. Lewis omits the case where it is the same person at two different times:

(4) I believe on 6 June 1977 'Today is Monday'. I believe on 7 June 1977 'Yesterday was Monday'. What is believed is the same.

I put it to you that this example does carry conviction. As argued above, if we deny it is the same belief, we are saying that the agent changed her mind. To re-cap the dialectic, all we need to refute Lewis is one example where a belief is less fine-grained than a self-locating property. That is, where the belief is retained while the self-locating property changes. And 4 seems to me a highly plausible example.

7   Perry's Solution

How does the two-dimensionalist handle the problem of the passage of time? What happens to the old belief that today is Monday as the clock strikes midnight? On a two-dimensional model, the belief has a content that is grasped by a role. The content of the belief has two components – the property of being Monday, and the day (d)

< Monday, d >.

This stays constant. It is eternally true and, we can assume, eternally believed. What changes is the role with which it is believed. On Monday, the day is grasped with the role expressed by 'today'. But roles must change as time passes in order to express the same content. Call the process by which roles change over time *mutation*[13]. Mutation is governed by simple rules such as 'the role expressed by *today is Monday* at t, mutates into the role expressed by *yesterday was Monday* at t + 1'. So the same content is grasped, firstly, with the role expressed by 'today is Monday' and secondly with the role expressed by 'yesterday was Monday'.

The crucial result we need to obtain dynamic beliefs is that 'Today is Monday' said on 6 June and 'Yesterday was Monday' said by the same person on 7 June express the same belief. This is the intuitive result that Lewis's theory denies. To guarantee this result we need a sufficient condition on belief identity. I offer the following:

If sentence 1 uttered at t1 and sentence 2 uttered at t2 express the same content, and the role expressed by sentence 1 has correctly[14] mutated into the role expressed by t2, then both sentences express the same belief.

---

[13] Mutation applies only when nothing that was uncertain is learnt. Think of someone watching the hands of a clock go round in a silent room, with the curtains closed and the phone unplugged. This may be an idealization that is never achieved, but idealization is a standard part of modelling agents (e.g. Bayesianism). Also, we sometimes learn when it is from a position of uncertainty, such as wondering what time it is and looking at your watch. This is not mutation but conditionalization (see below).

[14] By 'correctly' I mean to rule out cases in which the agent has lost track of time. For it is not obvious that Rip van Winkle, who has unknowingly slept for 20 years, retains the belief he

(I offer no necessary condition on sameness of belief. For example, I leave it open whether sentence 1 and sentence 2 express the same belief when uttered by two different people.)

So the two-dimensional theory allows that a belief can be retained over time, even though it is grasped with different roles, is expressed by different sentences, and corresponds to different self-locating properties. I think this ontology of dynamic beliefs is needed to avoid doing violence to our common-sense notions of belief retention and changing our minds, and this is worth the extra complexity of a two-dimensional theory.

Of course, there is nothing to stop someone adapting Lewis's theory to provide for belief retention in some way analogous to that suggested here (e.g. Schwarz (forthcoming); compare Chalmers' (forthcoming) 'enriched propositions'). But this will complicate Lewis's theory, which will undermine the motivation for his view (simplicity). My suggestion is that the resources already provided by two-dimensionalism can be appropriated to provide for belief retention. This completes my first argument against temporalism.

8 Belief Dynamics and Conditionalization

So far I have argued that the extra complexity of two-dimensionalism buys an ontology of dynamic beliefs. But if you think simplicity is very important and/or aren't concerned about dynamic beliefs, you won't be convinced. So I will try to sweeten the deal by arguing that the extra complexity of two-dimensionalism is independently motivated. Once we take confirmation theory into account, contents and roles map onto the two rules of belief update – *conditionalization* and *mutation*. This is evidence that contents and roles cut the world at its joints, or at least that the extra complexity is not gratuitous.

Standard confirmation theory – Bayesianism – admits only one rule of belief change: conditionalization[15]. This says that the degree of certainty in a belief after learning a piece of evidence should equal the earlier *conditional* degree of certainty in the belief, *given the evidence*. For example, suppose your conditional degree of belief that it rains give a thunderclap is 0.9. Now suppose that you do hear a thunderclap. Then your new degree of belief that it rains should be 0.9. Formally, if an agent has prior probabilities $P_0(H_i)$ at t0, and learns E and nothing else between t0 and t1, then her t1 probabilities should be $P_0(H_i|E)$, where $P(E) > 0$. Succinctly, $P_1(H_i) = P_0(H_i|E)$. This model of belief update is widely accepted by confirmation

---

expressed on going to bed with 'today is Monday' after waking up and holding the belief expressed by 'yesterday was Monday'. See Perry 1993.

[15] Jeffrey (1983) conditionalization has been included in the orthodoxy, but it doesn't affect the problem of *de se* beliefs so I will ignore it.

theorists.[16] But the model does not suffice once *de se* beliefs enter the picture.

Conditionalization only allows belief change when something the agent was uncertain about becomes certain. That is, there must be some evidence E that initially was less than certain, and then becomes certain. But clearly someone who learns nothing of which they were previously uncertain and who grasps a belief with the role expressed by 'today is Monday' should no longer grasp the belief with this role the next day. As we saw above, we need new rules – rules of mutation.

So we now have two rules of belief update – conditionalization and mutation. If we stick with Lewis's 'simple' theory, we have one object of belief that is subject to two types of change. On Perry's more 'complicated' theory, we have two components of belief and each has its rule of update: content is governed only by conditionalization; roles are governed by mutation[17]. So we see that Perry's two-dimensional theory is independently motivated. There are two rules of belief change, so there *should* be two components of belief.[18] Lewis's unified theory merely papers over a crack in the foundations and obscures an important distinction.[19]

(We often apply both rules at the same time of course. Someone watching a long boring film should simultaneously update the content ('the film is long and boring') and the role of his beliefs ('it is now past the time when this movie should have ended'). This kind of case is typical. My argument merely requires that these changes can be broken down into the two components of conditionalization and mutation.)

This completes my second argument against temporalism. The temporalist has to posit two different rules of update that apply to his single type of belief. The two-dimensionalist has two components of belief that match the two types of belief change. So any extra complication in her theory of belief is independently motivated. Before concluding I will mention an important objection that I discuss elsewhere.

The neat bifurcation I defend requires that content only changes by conditionalization. And this requires that mutation doesn't affect content. But many people believe that mutation does affect content. If so, the independent motivation for the two-dimensional theory would disappear; we

---

[16] See Howson and Urbach (1993) or Earman (1992) for influential texts.

[17] I cannot say that roles are governed *only* by mutation. After all, if you discover some new evidence E, the role component of your beliefs will change. But this is parasitic on the change in the content component.

[18] This coheres especially well with Chalmer's (2002) view that primary intensions are functions from centred worlds – which contain times - to truth-values, and secondary intensions are functions from uncentered worlds to truth-values.

[19] Indeed most writers on self-location and confirmation theory have assumed a Lewisian theory.

could then posit one type of belief that is governed by conditionalization, mutation and interactions of the two. Several philosophers have taken this approach and tried to develop such a theory (Halpern 2004, Meacham 2008, Titelbaum 2008, Kim 2009 Schwarz forthcoming), motivated mainly by the arguments of Elga (2000) (Sleeping Beauty) and Arntzenius (2003) (The Prisoner) which purport to show that an agent in certain circumstances should change their mind as time passes despite learning no new evidence that was previously uncertain.

I argue in Bradley (2011) that this is a mistake. I defend the claim that mutation cannot shift a rational agent's degree of belief in any content.[20] If I'm right, then the arguments that mutation can change content fail. So the view that contents should change only due to conditionalization, and not due to the passage of time, remains intact. So contents and roles have different, simple, rules of update. So the two-dimensional theory of belief that posits contents and roles is independently motivated. And whether or not my arguments are successful, there is I think an interesting connection here between philosophy of mind and formal epistemology (Titelbaum (this volume) comes to a similar conclusion).

Epistemologists have spent a great deal of energy arguing about how beliefs should and should not change when new evidence is learnt. Philosophers of mind have spent a great deal of energy arguing about how we should make sense of self-locating beliefs. But self-locating beliefs can change all by themselves, without any new evidence, and this creates a problem of belief retention for temporalism. I have suggested that an ontology of dynamic beliefs that fits naturally with the machinery of two-dimensionalism solves this problem.[21]

References

Arntzenius F. 2003. 'Some problems for Conditionalization and Reflection' *Journal of Philosophy* 100: 356-70

Bradley, D. J. 2011. 'Self-Location is No Problem for Conditionalization' *Synthese* 182: 393-411

Branquinho, J. 1999. 'The Problem of Cognitive Dynamics' *Grazer Philosophische Studien* 56:2-15.

Chalmers, D. 2002. The Components of Content. In *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.

Chalmers, D. 2011. 'Frege's Puzzle and the Objects of Credence', *Mind* 120:587-635.

---

[20] This is a logically stronger version of Titelbaum's (2008) Relevance-Limiting Thesis.

[21] I thank Dave Chalmers, Branden Fitelson, Chris Meacham, John Perry, Elliott Sober, Mike Titelbaum and Wolfgang Schwarz for helpful comments and discussion.

Earman, J. 1992. 'Bayes of Bust' Cambridge, Mass.: MIT Press

Elga, A. 2000. 'Self-Locating Belief and the Sleeping Beauty Problem', *Analysis.* 60:143-7

Evans, G. 1990. 'Understanding Demonstratives' in *Philosophy of language: the central topics* Ed.s Susana Nuccetelli, Gary Seay

Frege, G. 1892. 'Sinn und Bedeutung' *Zeitschrift fur Philosophie und philosophische Kritik* 100.

Frege, G. 1967. 'The Thought' in P.F. Strawson (Ed.) *Philosophical Logic.* Oxford: Oxford University Press.

Halpern, J. 2004. 'Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems', Proceedings of the Twentieth Conference on Uncertainty in AI, 226-34

Howson, C. & Urbach, P. 1993. *Scientific Reasoning: The Bayesian Approach* 2nd edn, Chicago: Open Court.

Jeffrey, R. C. 1983. The Logic of Decision (2nd Edition). Chicago IL, The University of Chicago Press

Kaplan, D. 1989. 'Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry and H. Wettstein (Eds.), *Themes from Kaplan* pp. 481-566. Oxford: Oxford University Press.

Kaplan, D. 1990. 'Thoughts on Demonstratives' in P. Yourgrau (Ed.) *Demonstratives.* New York: Oxford University Press.

Kripke, S. 2008. 'Frege's Theory of Sense and Reference: Some Exegetical Notes' *Theoria* 74: 181-218.

Lewis, D. 1979. 'Attitudes *De dicto* and *De se*' in D. Lewis (Ed.) *Philosophical Papers*, vol. 1. Oxford University Press, (1983).

Lewis, D. 1980a. 'A Subjectivist's Guide to Objective Chance' in Studies in Inductive Logic and Probability, Vol. 2

Lewis, D. 1980b. 'Index, Context and Content' In S. Kanger and S. Ohman (Eds.) *Philosophy and Grammar*. Dordrecht: Reidel.

Lewis, D. 1994. 'Chance and credence: Humean supervenience debugged' *Mind* 103: 473–90.

Lewis, D. 1998. *Papers in Philosophical Logic*. Cambridge: Cambridge University Press

Meacham, C. 2008. 'Sleeping Beauty and the Dynamics of *De Se* Belief' *Philosophical Studies*

Nuccetelli, S. and Seay, D. 2008. *Philosophy of language: the central topics* Rowman & Littlefield

Perry, J. 1977. 'Frege on Demonstratives' *Philosophical Review* (86) p.474-97

Perry, J. 1979. 'The Problem of the Essential Indexical' *Nous* (13) p.3-21

Perry, J. 1993. 'Rip van Winkle and Other Characters' in *The Problem of the Essential Indexical and Other Essays.* New York: Oxford University Presss.

Salmon, J. 1989. 'Tense and Singular Propositions' in J. Almog, J. Perry and H. Wettstein (Eds.), *Themes from Kaplan* pp. 331-92. Oxford: Oxford University Press.

Titelbaum, M. 2008. 'When are Indexical Beliefs Relevant?' *Philosophical Review* p. 555–605.