On the temporal characer of temporal experience, its scale non-invariance, and its small scale structure[1]

Rick Grush
Department of Philosophy 0119
University of California, San Diego
La Jolla, CA 92093

Abstract: The nature of temporal experience is typically explained in one of a small number of ways, most are versions of either retentionalism or extensionalism. After describing these, I make a distinction between two kinds of *temporal character* that could structure temporal experience: A-ish contents are those that present events as structured in past/present/future terms, and B-ish contents are those that present events as structured in earlier-than/later-than/simultaneous-with relations. There are a few exceptions, but most of the literature ignores this distinction, and silently assumes temporal experience is A-ish. I then argue that temporal character is not scale invariant, but rather that temporal experience is A-ish at larger scales (a few hundred milliseconds and above), and B-ish at smaller scales. I then point out that this scale non-invariance opens the possibility of hybrid views. I clarify (or modify, depending on how you want to frame it) my own view (Grush 2005, 2007) as a hybrid view, according to which temporal experience is B-ish at small scales – and at this scale my trajectory estimation model (TEM, a version of retentionalism) applies – but A-ish at larger scales, and at the larger scale my TEM does not apply. I then motivate this hybrid position by first defending it against arguments that have tried to show that the TEM is untenable. Since the hybrid view has TEM as its small-scale component, it must address this objection. I then put pressure on the main alternative account, extentionalism, by showing that its proponents have not adequately dealt with the problem of temporal illusions. The result is a new theory (perhaps characterizable as a refined version of my prior theory) motivated by i) explaining its virtues, ii) showing that objections to it can be met, and iii) showing that objections to its main competitors have not been met.

1 Introduction

An issue of current interest in the philosophy of mind is the nature of temporal experience, especially the relation between the temporal features of perceptual experience itself, and the temporal content grasped in that experience. While agreement is not universal, most players in the debate take as a starting point that the content of perceptual experience is not limited to snapshots of null duration, but rather concerns events spanning temporal intervals of some appreciable magnitude. One motivation for this is the observation that we seem to be able to *perceive* change, including motion. And since change can manifest only over temporal intervals, it seems as though the content of perception cannot be of the nature of a snapshot, but must rather comprehend a temporal interval non-zero magnitude.

What agreement there is on this point immediately dissolves to disagreement concerning what relation, if any, there is between the temporal content grasped in perceptual experience, and the temporal features of the experience itself. The main battle line is between approaches that try to get explanatory mileage from the fact that experience is (in a relevant sense) a process that evolves over time just as its accusatives do, and approaches that do not try to get mileage from that presumed fact. For now I will follow Dainton (2013) in calling the former *extensionalism*, and the latter *retentionalism*. There are other nomenclatures, and when the differences are relevant, I will discuss them.[2]

In Section 2 I will give a brief introduction to the topic, and describe extensionalism and retentionalism in more detail. I will then describe my own (Grush, 2005) trajectory estimation model (TEM), which is in some respects a type of retentionalist view. In Section 3 I will do two things. First, I will argue that temporal experience is not scale invariant, namely, that it may have one character at small scales and another at larger scales. I will clarify my own view as applying only at small scales, which is consistent with some other account – retentionalist or extentionalist – holding at larger scales. Though once scale noninvariance is seen as an option, other hybrid views become options as well.

In Section 4 I will address an objection raised by Barry Dainton (2013) against retentionalist views, the *problem of surplus content*. Though my own view differs from standard retentionalist accounts in some respects, it shares those features that render retentionalist accounts a potential target of Dainton's criticism. I will show how this objection does not pose a problem at the small scale.

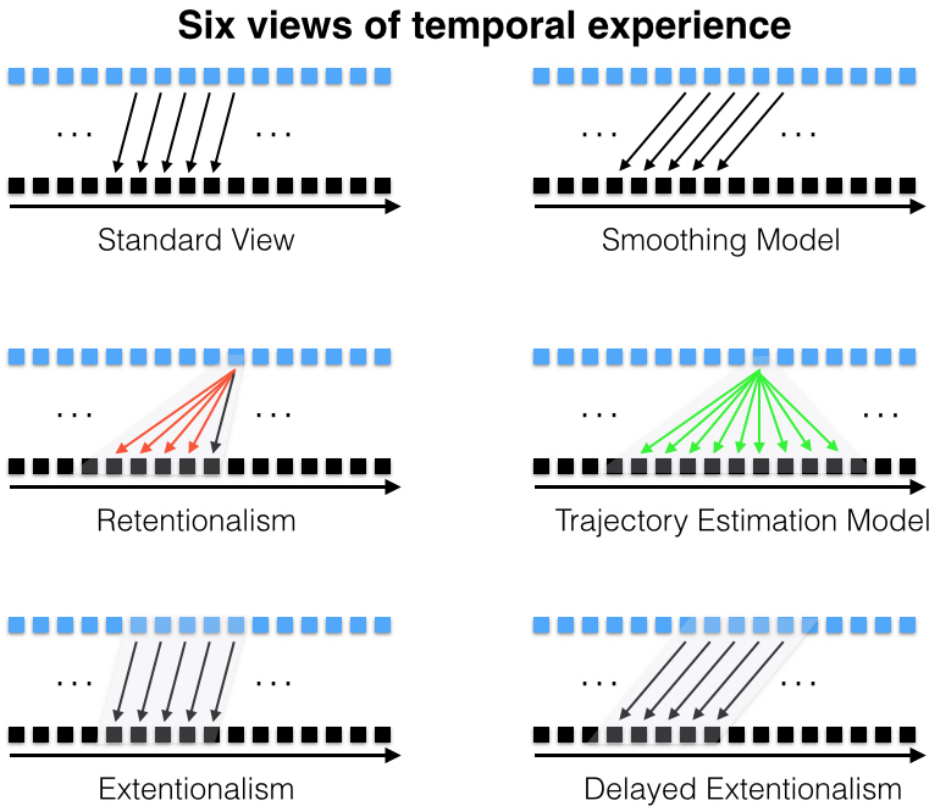
In Section 5 I discuss the prospects for extentionalism holding at small time scales, since the plausibility of my view depends at least in part on the plausibility of the alternatives. And the main alternative is extensionalism. So I will push an objection to extensionalism based on temporal illusions, and show that extensionalist positions have not adequately met it. In Section 6 I close with a general discussion.

[2] Lee (2014) distinguishes molecular from atomic approaches; Phillips (2014) distinguishes views based on whether or not they adopt what he terms the *transparency thesis*.

2 Overview of positions

There are six positions (some of which are variants of others) concerning the nature of temporal perception that I will discuss, though some get more airtime in the literature than others. There are others I won't discuss. The standard view is illustrated in Figure 1a. The bottom line represents time, the series of black squares just above the time line represents a sequence of real world states, and the series of blue squares above that represents a sequence of experiential states. The black arrows indicate the representation relation holding between experiential states and the environmental states they represent. Each experiential state at time t_i corresponds to an environmental state at t_i-d, where d is the *necessary* delay introduced by perceptual processing. The reason for emphasizing 'necessary' will be made clear in the next paragraph. On this view, the content of perceptual experience concerns only *states* of effectively null duration, though it does produce a sequence or stream (either discreet or continuous) of such states.

The smoothing model (Rao et al. 2001) is illustrated in Figure 1b. It is similar to the standard model in that it claims that the deliverances of perception are contents corresponding to effectively durationless states of the environment. The difference is that there is an *additional* delay, over and above the basic unavoidable processing delay d. The purpose of this additional delay, or *lag*, is to allow sensory information concerning environmental happenings after t_i to enable the system to construct a more accurate interpretation of what was happening at t_i. This is shown in the diagram by the fact that the arrows showing the correspondence between experiential states and environmental states point farther back than in the case of the standard model.



How does such a delay help? Supose that at time t_1 a baseball travels through the batter's box and at the same time the bat swings through in the opposite direction. At time t_1+d, there is information available to an onlooker's perceptual system suggesting that the baseball bat directly struck the baseball. But the perceptual system doesn't yet assume that that is what happened. It waits to see if further information will come in that might be relevant. At some point between t_1+d and t_1+d+l, (where l is the maximum lag for which the perceptual system will delay its processing) suppose that additional information does come in to the effect that the ball struck the catcher's mitt at t_i+l – perhaps the initial information included noise in the form of a random shadow flying in front of the batter's box that gave the appearance of a ball flying off the bat. This latter state (ball in catcher's mitt at t_1+l) is not consistent with the former (ball struck directly by bat at t_1), and so the earlier interpretation is relinquished. Because the perceptual system delayed its interpretation of what happened, it was able to take advantage of information concerning what happened after the event at issue to arrive at a better take on that event. This is obviously an advantage, but it comes at a cost, the delay. For starters,

it entails that, as Eagleman has described it, "We are living in the past", and by this it is meant that our conscious awareness is always of events that occurred in the past, indeed longer in the past than the minimum necessary delay for perceptual processing. This will be discussed in more detail in Section 4.4.

The expression 'retentionalism' has been used in a variety of ways in the literature. I will here describe it in a way that I think is at least close to center of mass of these views (Figure 1c). On this view, our perceptual experience *broadly conceived* of environmental *events* has two components: a perceptual component *narrowly conceived,* whose content is a punctate state, just as in the standard model. This is represented by the black arrow. But in addition to this, at the same time as this (narrowly conceived) perceptual state, there are a number of retentions, which are in some respects like memories of immediately prior environmental states. These are represented by the red arrows. See Dainton (2013) for more detail on this view.

So the *broadly conceived* perceptual state at t_i is a *complex of states* – a perception (narrowly conceived) plus retentions – hence the fact that all the arrows in 1c originate at one experiential state.[3] This complex provides to the subject *at* t_i*+d* a perceptual experience that spans a temporal interval, namely an interval from t_i*-l* to t_i, where *l* indicates the past-directed reach of retention, and *d* is the necessary delay required for perceptual processing. The content grasped by this perceptual complex concerns not environmental *states*, but rather temporally extended *processes* (aka *events*, aka *trajectories*).

My own model (Grush, 2005, 2007) is the *trajectory estimation model* (henceforth TEM, see 1d). The TEM is similar to retentionalism in that it maintains that at t_i*+d*, the subject has perceptual experience corresponding to a temporal interval. In addition, the TEM explicitly embraces a future-oriented component (corresponding to Husserlian (1991) *protention*), and so the represented interval spans t_i*-l* to t_i*+k* (*l*, $k \approx 100$ msec). One key difference between the TEM and standard retentionalism is that on the TEM, the perceptual state that constitutes the grasp of the experiential content is not an amalgam of multiple distinct kinds of states. Rather the TEM claims that there is a *single perceptual state* whose content corresponds to the entire temporal interval in question. Because of this, the TEM is perhaps more felicitously described as a type of what Lee (2014) calls an *atomic* view.

A second difference – at a minimum a difference of emphasis – is that according to the TEM, the content represented by these perceptual states is an *active construction, not a passive registration.* Successive perceptual states could have very different interpretations of the environmental goings-on over a given interval. For example, if a perceptual state at t_3 concerns the temporal interval from t_1 to t_5, and the next perceptual state at t_4 concerns the temporal interval from t_2 to t_6, they overlap, meaning that each will have an interpretation of what happened from t_2 to t_5. Importantly they may differ in their interpretation of what happened during that

[3] In Husserl's (1991) terminology the distinction was between retentions and primal impression, the latter being what I am calling the perceptual state narrowly conceived.

interval – that is, the latter perceptual state my revise the interpretation that was produced by the earlier perceptual state. Perceptions don't passively "sink back" into retention, as Husserl put it. Rather, the entire interval estimate is constructed anew each time (though the prior estimate is an important input to that construction, in most cases being something like the null hypothesis).

Traditionally, because retentions are often understood to be prior perceptual states that are simply passively *retained* by the overall system (rather than actively constructed), there is no obvious place for revision. This isn't to say that a retentionalist couldn't work that into their account. But it is an explicit part of the TEM.

The TEM provides the same perceptual benefit that smoothing provides, but without the cost of delaying. On the smoothing model, no percept is produced until after the delay period, and then a percept is produced that presumably will be, on average, more accurate because of the additional information accumulated during the delay. The TEM never delays the construction of a percept. But it does allow that as information comes in, how a prior event was interpreted can be *revised*. And so it will yield a better percept after the delay period if there is incoming information that warrants it – exactly as much better as, and no later than, the smoothing model gets with its delay.

Basic extensionalism (diagrammed in Figure 1e) maintains, roughly, that the process of temporal perception itself is a temporally extended process, and this fact plays a crucial explanatory role in understanding the temporal content of that experience. In the diagram this is represented by the fact that a sequence of the arrows are blocked together in light grey, indicating that they are all parts of an extended sequence of states. It is because my perceptual contact with an external event extends across the temporal interval through which that event unfolds that the *content* of my perception has the corresponding temporal features. That is, I can perceive an event that spans t_0 to t_1 – including perceiving motion – because my perceptual experience itself correspondingly spans the interval from t_0+d to t_1+d. For more detail, see Dainton (2013).

The smoothed extensionalist position is illustrated in 1f. It is basic extensionalism plus an additional delay, over and above d, added in order to produce better perceptual estimates. Experience still unfolds over time as with basic extensionalism, and this fact continues to explain the temporal contents grasped in experience, as with basic extensionalism. The difference is that that experience is delayed a bit. It is to basic extensionalism what the smoothing model is to the standard view. On this view, I experience an event spanning t_0 to t_1 because my perceptual experience itself correspondingly spans the interval from t_0+d+l to t_1+d+l.

A few remarks on these models before moving on. First, all views as far as I can tell can accommodate either a discreet or continuous interpretation. It is often easiest to describe these views in discrete terms, but this should be understood as a matter

of expressive convenience only. I don't believe that any of the points I discuss below hang directly on this.

Second, in some cases, most relevantly retentionalism and the TEM, the perceptual states are taken to manifest over a temporal interval that is significantly shorter than the temporal interval that those states represent. Sometimes this is described as the claim that the perceptual states are 'instantaneous'. But this is not a requirement. It is open to these views to maintain that these states can only manifest over some small interval – smaller than the interval grasped by the experience What is crucial is that the temporal features of the content of the experience are not explained by the temporal features of the experience itself.

3 Temporal character, scale noninvariance, and the possibility of hybrid views

There are two connected assumptions made by most of the discussion on these topics. The first cocnerns that I will call the temporal charatcter of temporal experience, and the second is that this temporal chartacter is scale invariant. What I mean by the *temporal* ***character*** *of perceptual content* is roughly whether the relevant temporal content is structured in terms of (i) past, present, future, or (ii) earlier than, simultaneous with, later than. These correspond to A-theoretic and B-theoretic temporal specifications in the literature on the metaphysics of time, respectively. The expressions *tensed*, *tenseless*, and *untensed* have been used by Hoerl (2009), and others, and although there is some overlap in the meaning he gives to these expressions, is it is not complete overlap, and I don't want to court confusion. So I will use the expressions *A-ish [temporal content]* and *B-ish [temporal content]* for this distinction.

At a *macro scale* our experiential contact with the evolving world appears to be A-ish. When I am watching a sprint in the Olympics, there is a point in the middle of a hundred meter race when I am perceiving the runners on the track, but am no longer perceiving their push off the blocks, nor yet perceiving their crossing the finish line. I may *remember* the former, and *anticipate* the latter, but there is a clear sense in which neither of those is part of my current *perceptual* experience of the runners at that time. In this way, my perceptual state at this scale marks out a *now* that I am perceiving as distinct from a past that I have perceived but am no longer perceiving, and a future that I will perceive, but am not now perceiving. And I think it is plausible that the same is true of slightly briefer time scales, such as a fly ball at the apex of its trajectory from home plate to the center field fence — a scale of about 2 seconds as opposed to the 10 or so for a 100 meter race.

But what about smaller time scales? It is natural to think that the temporal character of perceptual content is A-ish all the way down to instants of effectively zero duration. On this view, as the temporal scale decreases, temporal experience

continues to be characterized by a now flanked increasingly snugly by a past and future. At the limt we are left with a punctate now, which serves as a pointlike boundary between past and future, and which constitutes the temporal content of perception *per se*. Plus or minus some quibbles this is a natural way to understand the standard view and the smoothing model. On retentionalist views the narrowly-conceived perceptual component is often thought to provide the corresponding 'now' – Husserl called this "primal impression", and it corresponds to the "now-point".[4] Basic extensionalism and delayed extensionalism lend themselves to interpretation in scale-invariant terms along these same lines. An extensionalist might insist that there must be a span of experience of some minimal length before the result is *experience*, but whenever experience does result, the content would appear to be of an A-ish series of events regardless of the scale (though see Hoerl 2009).

But this scale invariance is not the only option. One scale non-invariant view would be that the temporal character of perceptual content is A-ish down to some bedrock limit, and beyond this limit there is no specifically temporal content. As we start slicing up temporal content, we eventually get to now-atoms of small but non-zero magnitude whose content is not sensibly divided further. The scale here is probably quite small, just beneath the briefest interval over which succession and simultaneity judgments can be made. That is, in the vicinity of 5-20 msec or so.

A more interesting scale non-invariant view would be that the temporal character of perceptual content is *A-ish* down to a certain point, and beyond that point it becomes *B-ish*. I can now clarify that my own theory is restricted in its application to the sub-200 msec scale, and the claim is that *at that scale* temporal content is presented in B-ish terms. No point in this interval is singled out as a *now* bracketed by a past and future. Rather, within this interval events are represented as standing in relations of *earlier than, simultaneous with*, and *later than*. But at scales greater than this, this B-ishly structured interval of 200 msec effectively becomes the bulky *now* of A-ish experience.

The idea that at large scales temporal experience is A-ish isn't terribly controversial, so I won't dwell on it. And the idea that there is something interesting about the scale of around 200 msec and less has been motivated in a couple of different ways. Dennett and Kinsbourne (1992) isolated something like this scale on grounds that it is the scale at which there stops being a well-defined finish-line for information being propagated through the processing mechanisms of the brain. I singled this interval out on the basis of theoretical considerations grounded in temporal illusions and representational momentum (Grush 2005).

But why think that at this sub-200 msec scale perceived events are presented in B-ish terms? One might expect an argument for this to be based on considerations similar to those Dennett and Kinsbourne raised concerning brief temporal scales such as this. In their article, Dennett and Kinsbourne leveraged the fact that the information processing machinery of the brain is physically extended and signal

[4] For example, he says "Primal impression has as its content that which the word "now" signifies" (Husserl 1991, §31, p. 66).

propagation speeds are finite in order to argue that at short temporal scales perceptual processing has features unlike those recognizable at a larger scale. Specifically, unlike the larger scale, at the smaller scale there is no identifiable point in time, or spatial finish line, that can be used to define a definitive *now*. I think this line of reasoning is basically correct. But Dennett and Kinsbourne were using those observation for other purposes. They were trying to argue against there being a single set of definitive facts of the matter concerning temporal relations at this small scale. This is not my topic. I am here not taking a stand on whether at this temporal scale experience is determinate or not. Rather, my argument is that at this scale whatever experience *does* present concerning the temporal features of the perceived situation – determinate or otherwise, in multiple drafts or single drafts – it presents in B-ish terms.

But here's a line of argument that is related to the Dennet and Kinsbourne point. Assessing the B-ish temporal relations between external events doesn't require an internal finish line. To see this consider the following toy example. There is a neuron in the brain that has two spatially separate retinal receptive fields, meaning that a stimulus presented in either of these fields results in an excitatory signal to that neuron. However, the magnitiude of that excitatory influence is enough to cause that neuron to fire only if stimuli are presented in both receptive fields at the same time. We have here a very simple circuit that detects *simultaneity*. It will fire when, and only when, two stimuli in the two different receptive fields are present at the same time. It need not keep track of time, or of what is *now* vs. *past*. There need be no special finish line. It simply encodes an assessment of B-ish simultaneity.

Of course one can easily imagine the mechanism failing to function in this way. If one of the pathways from the retina to the neuron was longer/slower than the other, the neuron would still fire when both excitatory signals reached the cell body at the same time, but because of the different transit times of the two pathways, the stimuli would have to be successive on the retina – with the stimulus in the delayed receptive field appearing first – in order for the signals to arrive at the cell at the same time to get the cell to fire. If this asymmetric delay were a rare occurrence (perhaps caused by transient biological changes in one of the axons targeting the cell) we would have intermittent perceptual error. On the other hand, if this happened consistently, then either through learning or through innate wiring the rest of the brain would stop treating that neuron's firing as a B-ish *simultaneity* signal, but rather as a B-ish *succession* or motion signal.[5]

The upshot is that at no point does this mechanism code for 'two stimuli in these receptive fields *now*' or 'one stimulus now and another just a moment ago'. The inputs to the mechanism take time. The stimuli were already past before the mechanism could do anything. Also, the signal leaving the mechanism itself takes

[5] Lee (2014) puts a mechanism such as this to service in a different way. Lee's point is to argue that the representing of temporally extended processes can happen in a temporally punctate way – succession coded in a single neuron firing. My point is rather about the character of the content represented, not the temporal features of the representing mechanism. Though they are different, the points are consistent.

time, and different amounts of time depending on which downstream processes want to use the output of that mechanism. If three downstream neural systems, all with their different transmission delays and times, want to make use of this simultaneity information (surely a huge underestimate), and if the information were coded in a tensed way, then we would need three different delay compensation mechanisms – as well as additional recalibration mechanisms for any communications between these downstream processors – in order to keep track of *now* and how far in the past the perceived events were. This is a lot of work to maintain the integrity of something one never really had to begin with: a 'these are both happening now' assessment.

But if the temporal content that is carried is a B-ish *simultaneity* assessment, then this signal maintains its relevance and correctness regardless of any delays, fixed or variable, that it may be subject to, through any of the pathways it traverses. It is temporally (and hence spatially) portable in a way that A-ish content isn't – precisely because its meaning does not hinge on a special reference point.

A tempting initial reaction to this line of reasoning might be something like the following: Maybe it's possible to get away with B-ish assessments such as this if all you are concerned with is sensation and perception, that is, making sense of the environment. But it needs to be remembered that the whole point of this is the control of action, that is, *engagement with* the environment. And when it comes to action, surely it is crucial to know what is happening *now* so that the organism can act appropriately. Just knowing that two events were simultaneous isn't enough. In order to act appropriately with respect to them, it also needs to be known whether they are happening *now*.

The temptation should be resisted. Exactly analogous points hold for sensorimotor coordination (at small scales) as hold for perception. What is important is that the baseball bat swings through the batter's box simultaneously with the ball's being at that same location. There is no *now* such that it is important that they both happen at that *now*, as opposed to some other *now*. Similarly, what is important is that my hand gets to the location below the vase earlier than (and maybe a specific temporal amount earlier than) the vase itself, so that I can catch it before it breaks on the floor. Whether any of this happens at any special time marked out as a *now* is irrelevant.

To put it another way, sensorimotor activity is a matter of seeing to it that some bodily actions are earlier than (perhaps by a specific amount), later than (perhaps by a specific amount), or simultaneous with, other relevant bodily and environmental events. And just as with sensory events, specifying motor actions in terms of tenseless relations relieves these and other systems from the burden of having to explicitly track delays or compensate for them. To be sure, it is crucial that transmission delays are accounted for. But the point of the accounting and compensation is to guarantee that things that should be simultaneous are simultaneous, and so forth. And this can be done, and done more easily, without having to go at it by way of coordinating anything with a master *now* assessment. And as a matter of fact, psychophysics experiments concerning temporal happenings

at small scales almost always ask subjects to make assessments of simultaneity, succession, and the like. Subjects are never asked to assess whether a stimulus was presented in the past vs. future.

A corollary of the implicit assumption about scale invariance is an assumption to the effect that with respect to temporal experience we are seeking after *the* ***one*** *correct account*. A denial of scale invariance opens up possibilities not explored in the literature. For example, the possibility that one might embrace an extensionalist explanation for the temporality of experience on a scale greater than a few hundred msec, but embrace TEM (or some retentionalist/atomist account) for the microscale of a couple hundred milliseconds and less. Another possibility would be a strictly atomist (TEM) account of what is happening at the small *B-ish* scale, together with a more traditional retentionalist view (distinguishing perception from retention as distinct psychological processes) at the larger *A-ish* scale.[6]

This leads to the last part of my view, which is that at a scale greater than 200msec, experience presents temporal relations A-ishly, and at that scale, a more standard retentionalist picture applies. In the rest of this paper, I will not argue for the part of the view concerning larger time scales. I will restrict my discussion to the small time scale, and try to argue that at least at this scale the TEM is viable, and that extensionalist accounts don't seem well-suited for application at the small scale. I will do this by pushing arguments against extensionalism (as applied to the sub-200 msec range) based on temporal illusions.

The arguments I will push against extensionalism only apply to that view at the shorter time scale, because temporal illusions are not in effect at larger time scales. And so I will say nothing against the possibility of a hybrid view that embraces the B-ish TEM at the sub-200 msec scale, and A-ish extensionalism at the larger scale.

4 The Problem of Surplus Content

An objection that Dainton has raised against retentionalist views in general, and which will apply to the TEM at the small scale, is what he calls the *problem of surplus content*. I will frame the problem in the terminology of my own view, but how it generalizes to retentionalist accounts generally should be clear.

On the TEM, the perceptual system grasps contents that concern a temporal interval of 200 msec or so. I have already explained that these perceptual states need not manifest instantaneously, but the idea is that they probably manifest in less than 200 msec. Just to get some numbers for ease of exposition, let's say that every 20 msec a new estimate of the 200 msec interval is produced. And this state persists for

[6] Lee (2014) briefly mentions the possibility of a view similar to the first sort I mention here, but dismisses it because he feels that his arguments against extensionalism/molecularism preclude any such hybrid account. Even if successful, Lee's argument wouldn't preclude a hybrid account that didn't make use of extensionalism/molecularism, such as the second possibility I discuss.

20 msec., at which point it is replaced by its sucessor. If this is so, then over a course of one second of objective clock time, the perceptual system will produce 50 separate trajectory estimates, each constituting a grasp of 200 msec. of temporal content. The math says that on this account the perceptual system is producing a total of ten seconds worth of content every second! And this is Dainton's *reductio* conclusion. Any view according to which perceptual episodes manifest over an interval shorter than the interval their content concerns will be a target of this objection. As Dainton puts it (using different numbers, but the problem is the same):

> Let us suppose … that the apparent duration of the specious present is approximately one second. According to the Retentionalist we are experiencing this much apparent change at each and every instant. It follows that over the course of one second vastly more than one second of change features in our experience: an infinite quantity if specious presents are densely packed, but even if we suppose there are only a hundred specious presents packed into each second, there will be more than a minute of experienced change per second. But where is this additional content? Over a one second period, aren't we typically aware of just one second of change? An account which generates significant quantities of surplus-to-requirements phenomenal content may not be incoherent, but it has a severe plausibility problem, to say the least. (p. 377)

According to the TEM, what is happening at each perceptual episode is the grasping of 200 msec worth of temporality in B-ish terms. One grasps that there were *n* events, standing in relations of earlier-than, simultaneous-with, and later-than, with each other, such that the temporal distance between the earliest and the latest is about 200 msec. It is a small-scale version of the view that Spinoza's God has on the history of the universe *sub specie aeternitatis*. And as with Spinoza's God, the appreciation of the temporality of this interval does not require a special or perspectival now-point that progresses through the interval.

It is worth saying a bit about these two different ways – the *now-bound* and the *divine* – of appreciating temporality. Consider a current process that we are experiencing, such as the melting of the polar ice caps. For us finite beings, our experience of temporality involves a now-point to which our experience is tied, and our experience of temporally extended events such as this involves the progression of this now-point through the relevant interval – a progression that takes as much time as the length of the experienced interval itself. We are experiencing the melting of the polar ice caps, precisely because we occupy a now-point that is progressing through the same temporal interval as this event. Things are different for Spinoza's God. He takes in the event *at a glance*. God is not *experiencing* the melting. It would be much better (though still not entirely felicitous) to say that God simply experiences it (perhaps better: *knows it*) – along with everything else in the history of the universe – all 'at once'.

Of course, *if* in order to experience a 200 msec interval one needs to A-ishly progress through it by surfing the cusp of a now-point, a problem of surplus content will be easy to generate for any theory that maintains that this can be done in substantially less than 200 msec. And especially if the theory claims that this is done

50 times a second! We would have to engage in ten seconds' worth of now-point surfing every second.

But what should we say of a being that grasps the 200 msec interval B-ishly, all at a glance? And what should we say if there are successive 'glances' at these intervals that partially overlap? How do we *sum* successive overlapping takes on the same thing? There are two ways to do the accounting. Perhaps an analogy will help.

When Zamboni drivers resurface the ice in a rink, they drive in ovals which overlap. The reason for the overlap is that it is important that all of the ice be resurfaced, but driving precision isn't tight enough to guarantee that subsequent swaths of resurfacing will exactly abut such that no unresurfaced ice remains between swaths without overlap. As a result, some parts of the rink are resurfaced more than once. To get some numbers, let's assume that on each loop, there is a 20% overlap with the previous swath of resurfacing.

One could ask the following question: how much ice was resurfaced during a complete Zamboniing session? A standard hockey rink is about 1580 square meters. So in one very straight-forward sense, 1580 square meters of ice was resurfaced. On the other hand, we have a two-meter-wide resurfacing blade, oriented perpendicularly to its direction of motion, that travelled for 948 meters, yielding 1896 square meters of resurfaced ice.

Now clearly we have not produced a reductio argument against the possibility of Zamboniing, by demonstrating that if a rink were Zambonied, then 1580 = 1896. Rather, we've unearthed a harmless ambiguity in how we might sum overlapping magnitudes. On the one hand, we can reckon the summable magnitude with respect to the Zambonied entity, i.e. the ice rink. On the other hand, we can recon the summable magnitude with respect to the thing doing the Zamboniing. In the former case we have 1580 square meters of ice in the rink, and all and only *that* ice was resurfaced. In the latter case 1896 square meters of resurfacing was done.

So the question is, according to the TEM, how much experiencing takes place per second? Here is one answer: On the TEM, the experienc*ing* occurs in 20msec blocks (say, I'll use the same numbers assumed in the previous section), and there are 50 of those per second. So on the TEM there is 1000 msec of experienc*ing* per second. The math checks out. No surplus experienc*ing* is going on.

Here is another answer: On the TEM, since each episode constitutes a grasp of 200 msec of temporal content, and there are 50 of these per second, during each second ten second's worth of temporal content is grasped.

But this by itself doesn't generate a problem. True, we have multiple overlapping graspings of this one seconds' worth of temporal content. But we can get from that to problematic surplus content, as far as I can tell, in one of two routes, neither of which yields a sustianable objection to the TEM proponent.

First, one could maintain that the only way to experience 200 msec of temporal content would be by riding an experiential now-point through those 200 msec. I agree that this would immeuately generate a problem. But because the TEM claims that these grasped contents are B-ish, this does not apply. God's experience of the lifetime of the universe takes no time; and our experience of the 200 msec. is similarly atemporal – *sub specie short-intervalatis*, so to speak. The grasping of the 200 msec. of temporal content happens at a glance, and does not require traversing 200 msec. of perceptual now-point.

The other route would be to insist that the only correct way to sum the contents is by double-counting the overlaps. But I can see no reason why the sum must proceed this way. And if this is supposed to be a threat to TEM, surely there needs to be an argument as to why we have to count in this way. We *can* do it that way if one insists, but as far as I can see the problem that would result is no more troubling than the *problem of surplus ice*.

Indeed, we could generate an exactly analogous problem for binocular vision. When looking at a one meter oak board, if my right eye sees one meter of oak, and my left eye also sees one meter of oak, then I could, in some sense, claim that I am seeing two meters of oak. While I could do this, why would I want to? It seems that the right thing to say is that I see one meter of oak, even though my seeing of it is redundant. But the redundance shouldn't be used as a reason to sum the magnitudes separately. What I am seeing (redundantly) is the one meter board. Similarly, what I am percieiving is the on second of change. I am perceiving it reduncdantly, in that each phase is perceived by me on multiple occassions.

And so as I have shown, on the TEM there is one second of experiencing per second. And there is also one second's worth of change that is (redundantly) experienced every second. The math works out. As far as I can tell, the only way to generate a problem is either to insist that experience must invovle surfing a now-point (which the TEM denies, because it posits B-ish contents at the small scale), or insist on summing magnitudes in a way that would also, and quite objectionably, have me seeing two meters of oak every time I look at a one meter oak board with both eyes open. Neither of these strikes me as high percentage.

5 Temporal Illusions and Extensionalism

Part of the plausibility of the TEM rests on objections that have been leveled against views of its type. Another part depends on the extent to which competitors are seen as plausible. My goal in this section is to show that, at least as applied to the sub-200 msec scale, extensionalist models face serious challenges. I will be pushing the argument from temporal illusion. Extensionalists are familiar with this general sort of objection, as I raised them (Grush 2007, 2008) against Dainton's view. I will expand on the argument from temporal illusion by showing not only that

it has more force against Dainton's view than he recognizes, but also showing how it presents a serious challenge to the views of Hoerl (2009) and Phillips (2014) as well.

The basic argument can be summarized fairly quickly. Consider a temporal interval from t_1 to t_3 during which a subject is looking at a blank computer screen in a dark room. At t_1 a light flashes at location A on the computer monitor, and at t_3 (a hundred milliseconds or so after t_1) a second light flashes at a location to the right of the first flash, location C. This is just the standard apparent motion stimulus condition.

As perceptual processing progresses, there will be a point at which the information processing responsible for the experience of the first flash is complete, this will be the time that the flash actually appears plus the required processing time d. So at time t_1+d, the subject's perceptual system will have received and had time to process information from the first flash. But let's consider the time t_2+d, and ask ourselves, what percept is the perceptual system producing at this time? At t_2, the first dot has flashed, and given way to a black screen. The amount of time it takes for information to be processed about the flash on the screen and the return to black has elapsed by t_2+d. But the second dot has not yet appeared. It would be natural to say that what is experienced at t_2+d is the empty black screen. And perhaps this black screen is even experienced *as* having just supplanted an isolated flash, though this is not crucial for the argument.

When we get to t_3, the second dot flashes, and at t_3+d the time it takes for information about the second flash to make its influence on perceptual processing has elapsed. And we know that in this situation, people perceive the event *not* as two isolated flashes – at location A at t_1 and at location C at t_3 – but rather as a dot moving continuously from A through B to C.

But – and here is the puzzle – it seems that in order to see the scene as one of continuous motion from A through B and on to C, the perceptual system must represent the first flash *not* as followed by an empty black screen at t_2, but as followed by a stimulus moving away from A through B towards C, and in particular with the stimulus at B at t_2. So it would seem that experience is presenting both i) the first flash followed by a completely black screen, and ii) an initial stimulus that starts at A and continues through B. These two possible contents are obviously incompatible.

Here is one way to handle the problem: the perceptual system produced a representation of the black screen at t_2+d but this representation was soon replaced by a different interpretation (of a moving dot) without memory of the first representation. But *prima facie* (see below for more) the extensionalist cannot make use of this reply. It is the actual progression of my perceptual experience, between t_1+d and t_3+d, that explains my experience of that temporal interval, according to the extensionalist. So whatever I *actually* experienced at t_2+d is part of that experience.

Moreover, if over-writing could solve the problem, then the extensionalist would be admitting that in this case, a current perception at t_3+d, plus a currently produced and grasped (false) memory (at t_3+d) of there having been a moving dot at t_2+d would explain our experience of the motion at t_3+d. But this possibility just is standard *retentionalism*., the idea that the experience of a temporal interval is an amalgam of a current percept plus memories (retentions) of what just preceded. The extensionalist must avoid appeal to memories (false or veridical) as sufficient, when combined with a current perception, to explain the experience of motion on pain of becoming a closet retentionalist. The extentionalist must appeal only to the *actual* temporally extended stream of experience from *at least* t_1+d through t_3+d.

With overwriting ruled out, there have been three extensionalist proposals concerning how to deal with temporal illusions. The first is delayed extensionalism – the option embraced by (Dainton 2008). Second, Hoerl (2009) argues that we can attribute to the subject at t_2+d an experiential content that is 'neutral', with the neutrality resolved some time after t_3+d. Finally Phillips (2014) maintains that we can attribute to experience a content at t_2+d that is determinate (not neutral, *pace* Hoerl) and not delayed (*pace* Dainton), but which depends on facts concerning what happens after t_2+d. Note that the TEM avoids this problem because it explicitly embraces the ability to over-write past experiences.

5.1 Dainton's view: delayed extensionalism

Dainton has claimed that the problem can be solved by recognizing a delay in the processing. As Dainton puts it:

> We can agree that experiences cannot have inconsistent contents. But it would be wrong to suppose the Extensional theorist has no option but to interpret such cases in this way. Grush seems to be assuming that the contents featuring in Extensional specious presents reflect their environmental causes in an immediate and entirely unmediated manner. But there is no need for Extensionalists to embrace this view of the perceptual process. It is arguably more plausible to construe perceptual contents as representations that are generated in the brain only after a good deal of processing. This processing makes for a delay—50-100 msec, say—but our brains put this to good use: they try to work out a single, coherent version of events on the basis of the fragmentary and (at times) conflicting data available to them. Only this 'final draft', as it were, reaches consciousness. Hence in the phi case, our perceptual systems reach the (in fact erroneous) conclusion that A is in fact a moving light, and this is the only way in which it features in our experience. While the initial solitary, static A-flash may well register in our perceptual systems, it does so only at a pre-conscious level. Since this flash in this form is not experienced, the problem of inconsistent perceptual contents does not arise. (Dainton 2008, pp. 381-2)

I agree that this will get around the problem of contradictory contents. But this solution comes at a higher cost than Dainton realizes.

5.1.1 The cost of delay

Before beginning, a clarificatory point needs to be made. In order for a delay to solve the problem it must be a delay *over and above* the minimal delay required for perceptual processing. Whatever minimal delay the sensory information from the earlier event is subject to (we have been calling this d) applies equally to information from the later event. Supposing the later event (e.g., the second flash) occurs at t_3, then this information won't be in a position to make any postdictive difference to anything until t_3+d at the earliest. This is not nearly as innocent as the claim, made by Dainton after pointing out that all perceptual processing unavoidably takes time, that the perceptual system "puts this delay to good use" suggests. That phrase makes it sound as though the time period required for processing the *initial* stimulus information can serve as the delay period required for retrodictive processing – that the brain is just multitasking for that same delay period. This is not the case, however. That is, if, as Dainton suggests, "… the initial solitary, static A-flash may well register in our perceptual systems, it does so only at a pre-conscious level" it needs to be exmphasized that on this proposal, the subject can only become aware of the A-flash at t_3+d. The subject's experience is delayed, and Eagleman's claim that we are living in the past is embraced.

Next, it needs to be emphasized that the perceptual system does not ever know in advance if it is in a situation such that data received after an event will be relevant to the perceptual interpretation of that event. In order to know this, it would have to look into the future and know, for instance, that a second flash is coming up, and so processing should be delayed concering this first flash to take account of it. But obviously if it could know this, then it could use this clairvoyance to solve the problem directly without delay. This means that on any view that posts a delay to address temporal illusions, the perceptual system must be *continually* maintaining a perceptual delay, in *every modality*, *every waking moment*, just for those relatively few situations in which a delay will afford a better interpretation.

It would be difficult to quantify the costs and benefits here. By my lights, the introduction of an additional 80-100 msec delay, in all modalities, 24/7, seems like a fairly hefty price. And while there is the benefit of better interpretations from time to time, it's tough to put a number on how much better, how often, and how useful these better interpretations are. My own sense is that this benefit it is much smaller than the cost. (This is one reason I endorse the TEM, which has *all of the benefit* of retrodiction but *none of the cost*. The cost-benefit accounting is much more straightforward in those conditions.) But I recognize that intuitions on this could vary widely.

Why are ubiquitous perceptual processing delays a bad thing? We believe that we execute actions based on our conscious perceptual experience. And much of our

daily behavior is such that normal performance involves acting on the basis of sensory input within latencies that don't seem to allow for this sort of delay – when playing many sports, for example. We would seem be forced to admit that conscious perceptual experience is, in a way, epiphenomenal, at least with respect to events at a fast time scale. The defender of delayed extensionalism (or the smoothing model) might reply by saying that motor control in this short time frame could be accomplished by *subconscious processes* that make use of perceptual representations that are not delayed. The delay applies only to something like what gets passed along to conscious experience.

But this would be an odd response to make, since if there's any point at all to positing a purposeful delay in order to arrive at a superior interpretation of environmental happenings, it is that *this superior interpretation will result in better behavioral guidance*. But this response is admitting that that behavior is governed by something (unconscious processes) that *isn't getting this benefit*; and moreover that the thing that is getting this benefit (conscious experience) can't do anything with the benefit because it is delayed out of the relevant causal loop. If the delay isn't aiding the mechanism that could benefit from it, but is only aiding a mechanism that can't benefit from it, then what in heaven's name is the point?

5.1.2 The requirements of retrodictive processing

The TEM does not face the costs just itemized, and so all else being equal it would seem like it should be the preferred approach. But the proponent of delayed extensionalism will claim that all else is not equal, that the TEM is computationally expensive and requires a lot of fancy processing. Enough, perhaps, that extensionalism might still on balance be the better choice.

It is true that when I have presented the TEM (Grush 2005, 2007) there has been a good deal of math in order to specify the processing requirements. By contrast, all that the smoothing model and delayed extensionalism do is to 'delay processing to be able to take account of information from events that occur after the perceived event'. And that sounds easy enough. But appearances are deceiving. Failing to be explicit about (or cognizant of) the required complexity of a process does not render that process any less complex. And in fact, smoothing and delayed extensionalism require all of the information processing machinery that is needed by the TEM. They just put this machinery to far less efficient use.

Let me explain. Perception, as standardly understood, is a process that starts with raw sensory information transduced (usually) at the periphery, and results in a percept that is as of the environment or some object in, or feature of, it. To put it in schematic terms that are intended to be uncontroversial, perception requires that there be some knowledge or expectations about what the environment is like and what sorts of things generally happen in the environment, and these are used to

interpret sensory input. Stimulation of the visual sensory surfaces is limited to two-dimensional patterns of light, but what we perceptually experience are chairs and tractors and the dog's pursuit of the ball – things that are not transduced at the sensory surface, but rather supplied by the perceptual system.

Apparent motion is a reflection of these expectations: the perceptual system evidently thinks that a single moving object is more likely to be the cause of the sensory information than two distinct but successive bright stimuli in close spatial and temporal proximity. Most perceptual illusions are the result of such top-down expectations in this way. What makes this sort of thing necessary is the relative paucity of peripheral sensory information and its noisiness. In signal processing terms, this process is often called *filtering*. It can sound fancy, but all it means is a process by which expectations are used to interpret sensory signals – roughly, those things that better fit expectations are more likely to be signal, and those which do not are more likely to be noise. We see the Neckar cube as a cube because our perceptual system takes it that that sort of 3D object is a more likely cause of that stimulus pattern than 12 co-planar line segments.

Now let's turn to *retrodiction,* aka *smoothing*. This is a process by which information gathered after some event is used to help arrive at a better idea of what that event was. But how can information from t_2 help us to know what was happening at t_1? The only way is expectations about how the environment (or perceived/represented) system is likely to evolve over time. Many of these expectations are so banal that it is easy to overlook their necessity and even existence. When I open my eyes at t_2 to see the bowling ball rolling towards the pins, I can know that it was very recently (t_1) thrown by the person standing at the top of the lane because I have expectations to the effect that i) bowling balls don't pop into existence randomly; ii) they tend to travel in straight lines. And accordingly one can get an idea of where they came from based on knowledge of a chunk of their current path. That is, one can use current information to help determine what happened in the past, but only because of expectations about the kinds of things that typiucally occur. Without a lot of expectations about how the perceived entities typically behave, retrodiction would be impossible.

So in order to do retrodiction at t_2 concerning what happened at t_1, I need to have two things: first, an idea of what the current situation is at t_2; and second an idea of how this sort of system behaves over time. And then I use the second piece of information to, so to speak, *backtrack* from the current situation to a guess as to what the prior situation was. I then use both i) the percept that was produced at t_1, and ii) the back-tracked guess, produced at t_2, concerning what must have happened at t_1, to decide on the best interpretation of what was happening at t_1. This might require over-writing the original t_1 estimate, maybe averaging them, or going with the original one despite what later information suggested. Which route is taken depends on a number of factors. The points I want to emphasize now are that the backtracking to the better interpretation of what was happening at t_1 i) starts with an interpretation of the current situation at t_2, not just the raw unprocessed information caused by the situation at t_2; and ii) it requires expectations about what sorts of events are to be expected in the environment.

Next, note that *these very same ingredients* – a current perceptual state indicating the current situation and expectations about how the things typically behave – are sufficient to generate predictions of future states of the environment. This is self-explanatory.

To summarize so far: the smoothing model and delayed extensionalism require i) the ability to produce percepts of the current situation, and ii) expectations about how events typically unfold. They use these ingredients to actually produce an estimate of the current situation as well as an estimate of prior events that lead up to the current situation. They could easily use these ingredients to produce predictions of future stages of currently perceived processes, but the models don't make use of this ability.

Now let's look at the TEM. It maintains that the perceptual system constructs estimates of the behavior, over a period of time from a brief interval in the past to a brief interval into the future, of the perceived environment (here 'past' and 'future' are descriptions of the extent of the interval with reference to the time that the perceptual state is produced, it is not a characterization of the character of the temporal content of that state, of course). What this requires is an ability to generate estimates of current states of the environment (perception), expectations about how the environment typically behaves to produce both estimates of the past states of the environment, and also estimates of its future states. Put them all together, and you have a *trajectory estimate*. That is, the TEM and the smoothing model/delayed extensionalism both require the same computational power, and both use that power to produce representations of the most current environmental state as well as representations of what prior events resulted in the currently perceived events. Only the TEM maintains that the machinery is also used to produce predictions of future stages of perceived happenings.

In what do they differ, then? The TEM maintains that the entire trajectory estimate is part of conscious experience, and also that the entire estimate is available to guide behavior, even by sub- or pre-conscious processes. Extensionalism/smoothing model place a limit either on what is presented to conscious experience, or what is available to guide behavior, or both. I have remarks above that indicate the main problems with this (in 5.1.1). The point is that despite the fact that the proponents' desription of the process of benefitting from delay sounds easy and computationally cheap, it isn't any less expensive than the TEM.

5.2 Hoerl's view: non-delayed 'neutral' content

The question concerns what the subject perceives after the first flash, but before the second flash (if there is a second flash). The challenge was that it seems that the subject must experience the isolated flash at t_2+d, but then this is inconsistent with

what we know the subject will perceive if the second flash is presented at t_3. There are two components to Hoerl's view. Notice that in the discussion below of Hoerl's view we will be switchng the time indicies of the flashes from t_1 and t_3 with a blank screen at t_2, to t_4 and t_6 with a blank screen at t_5. The reason for adding 3 to each index will be apparent shortly. Note also that Hoerl's expression for extentionalism is 'molecularism', for current purposes we can treat them as synonymous. The first component of Hoerl's response is this:

> I think the most promising strategy for the molecularist to adopt, in response to this objection, is to insist that experiences can't be sliced arbitrarily finely. In other words, there is a limit to the fineness of grain, as it were, of the temporal phenomena we can make out perceptually. … If there are such thresholds, though, the molecularist does not need to accept the crucial idea behind [this] criticism, that the overall temporal experience of what happens between t_4 and t_6, as conceived by the molecularist, must have as a proper part a further temporal experience, *viz.,* of what happens between t_4 and t_5 . (Hoerl 2009, p. 10)

The idea here is that if there is a minimal temporal duration of experience that can count as experience, then perhaps the extansionalist (aka molecularist) can deny that it makes sense for us to ask after what is happening the time between flashes. What ever is happening in thiss short a time-frame is too short to count as an experience.

To this two responses could be made. First, supposing this is true, then it would seem to be a version of Dainton's delayed extentionalism, but just posited for different reasons: at t_5 (the time between flashes) not enough time has elapsed for the experience to count as an experience. But by t_6 it will.

The second reply is one that Hoerl credits to an anonymous referee. It is this: let's not consider the interval from t_4 to t_5, but rather from t_1 to t_5. We can now ask: what was the subject experiencing at t_5? And now Hoerl can't reply that the interval of experience isn't large enough to count as an experience, since we can make that interval as large as we please. (This is why we switched index numbering.) To this Hoerl replies:

> In my view, this objection relies on the same problematic step …. The only thing that could motivate the claim that this experience must be as of a period with a stationary stimulus occurring at the end is the idea that it has as a proper part a further temporal experience, of what happens between t_4 and t_5. (Hoerl 2009, p. 11)

Suppose I am watching a parade with seven floats, and each float goes past me at t_1, t_2, …, t_7. And let us suppose that any three successive time steps is sufficiently long to count as an experience. Whatever the minimum experience length is, just set the indices so that that much time is covered in any three successive time steps. So t_1 - t_3 is long enough for an experience, as is t_4 - t_6. Here are some things that strike me as true: i) my experience spanning t_1 - t_3 does not include experience of the 4th float; ii) my experience spanning t_1 - t_4 does include experience of the 4th float; my

experience spanning t_4 - t_6 does include experience of the 4th float; but my experience spanning t_5 - t_7 does not include that experience.

What this suggests is the following: consider an extended chunk of my experience that includes t_4. It something about my experience at t_4 that is selectively responsible for the fact that my experience over the longer interval containing t_4 includes experience of the 4th float. (This is a really fancy way of stating something that is pretty plausible pre-thyeoretically.) Note that this principle does not require the further claim that the chunk of experiencing that is happening at t_4 could, by itself, count as an experience. Identifying which temporal parts of an experience might be selectively responsible for certain contents being present in an extended chunk of experience does not require one to think that that part could, by itself, count as an experience. And it seems to me that the weaker claim is all that the anonymous referee's point requires. And this does not require the assumption that Hoerl finds problematic.

Now to what I see as Hoerl's main response:

> The segment of experience might be such that it does not distinguish between a stationary stimulus occurring at its end, and the onset of a movement of a stimulus occurring at its end. In other words, the molecularist can argue that any segment of experience occurring before $[t_3+d]$ will be neutral as to whether what is experienced at its end is a stationary stimulus or the onset of a movement of a stimulus; that neutrality is removed only at $[t_3+d]$. (Hoerl 2009, p. 11)

Note that I have switched the indices back to the ones used in the prior discussion. What can it mean to say that experience is 'neutral' or 'does not distinguish between' i) an isolated stationary stimulus that elapsed leaving only blackness and ii) a bright stimulus moving across the screen? It seems that there are two ways to go. First, it might mean that there is *no substantive content* at all. Second, it might mean that there is substantive content, but that it is subject to more than one interpretation, namely, the two interpretations that it is neutral between. An example is the visual appearance of a Neckar cube, which has a content, but is ambiguous because this content could be interpreted in a three ways (perhaps more, but there are three that are most common). Perhaps with some measure of expressive license this could be described as the 12 co-planar line segments being "neutral" between the two possible cube percepts.

It is clear that Hoerl has in mind something like the second interpretation. In any case, if he had in mind something like the first, then the position would collapse to Dainton's delayed extensionalism since there would be no content until after the second flash, and my remarks concerning delayed extensionalism would apply. On the other hand, I don't think the position interpreted as positing neutrality understood as *ambiguity* is tenable. In order to be an adequate way of conceiving of things, the neutrality will have to be between all possible eventual perceptual contents. Hoerl's discussion focused on two possibilities — an isolated flash and a dot moving from A to C — undoubtedly because he was responding to my discussion in (Grush 2008), and these were the two possibilities I focused on. In my case, the two

options were enough for the point I was trying to make. But they are not enough for Hoerl's point.

What are the possible interpretations the perceptual system might land on, at t_3+d, concerning what is happening just after the first flash, at t_2? We have seen two: a black screen that just displaced an image of a flashing dot, and a moving dot at location B that just left location A and is headed for C. But the description of the second possibility is deceptive. The second dot could flash in a location *any direction* relative to the first. The motion could be upward, leftward, rightward or downward. Indeed, there are a very large number of possible discernible directions. How many precisely is a question for psychophysics, but essentially as many motion trajectories as there are discriminable locations for the terminal flash. Dozens at a minimum. Perhaps hundreds.

And the possibilities expand further when we recognize that the color of the second flash will also influence the interpretation of what happens along the trajectory, since when the two circles differ in color, the illusory moving circle is seen as changing color before the terminus. If we multiply the discernible locations by discernible colors, we see that the ambiguity of the post-flash percept is high indeed, into the thousands at least. The percept must be neutral between motion in any direction, and a transition to any color.

And it has also been found that the shape of the second flash affects what is seen along the trajectory, with an initial circular stimulus morphing continuously into the square shape of the terminal stimulus, for example (Kolers and von Grünau, 1976). It is unclear what the limits are for this sort of transition (will a disc morph continuously into the shape of a six-legged elephant?). We may be into six figures by now, perhaps much higher. Worse still, if the stimulus presented at t_3 is a mask, the percept could be one of no initial flashing disc at all!

I'm unsure what to make of the suggestion that my perceptual experience at t_2+d concerning what is happening at t_2 is a) substantive in the sense that it is not just a lack of content waiting out a delay period as per Dainton's position, but b) neutral between there being an isolated flashing disk, there not being a flashing disk at all, there being a flashing green disc that is starting to morph into the shape of a periwinkle banana while moving down and slightly to the left, and so forth for several hundred thousand (at least) trajectories with the flashing disk as an initiator. Even with two possibilities my grip on what this was supposed to mean was shaky. But now I fear I've no traction on the suggestion whatsoever. It is also puzzling what possible use it could be for the perceptual system to go through the trouble of producing that sort of strange neutral percept. Especially when the equally or more useful option of producing no percept at all — that is, *delaying* — is right there.

5.3 Phillips' view: 'future-dependent' experience

Ian Phillips (2014) makes the following response in defense of extensionalism (note that his example is the cutaneous rabbit, and not apparent motion):

> … if a subject's experience at a time is logically dependent on facts about experience at later times, we cannot say what a subject is experiencing immediately after the second pulse without taking into account facts about her later experience, and in particular whether she goes on to experience a sixth pulse or not. In consequence, a subject may have different experiences of the second pulse's location – without need for delay or over-writing – because such experiences constitutively depend on facts about later experience. (Phillips, 2014, p. 136.)

To summarize: There is experience at t_2 (the second pulse), and the question is whether it has feature X or feature Y — this could be the experience's being as of an isolated flash vs. the initial stage of a moving dot, or it could be a tap felt at the wrist vs a tap felt a few centimeters proximal to the wrist. We are told that this might depend (logically) upon something that occurs at t_3. For example, if E_1 occurs at t_3, then experience will have feature X, but if E_2 occurs at t_3, then experience will have feature Y.

5.3.1 Marty McFly and game-winning shots

The rarified metaphysical air of various sorts of dependence (logical and otherwise) is a bit much for me. I will rather pitch my discussion in terms of a distinction between two exhaustive (I believe) sorts of dependence, namely the sort of dependence that yields (1) below, and the sort that yields (2):

1. The feature that is determined at t_2, by the fact at t_3, is such that it is able to make a causal difference to events that occur *between* t_2 and t_3.

2. The feature that is determined at t_2, by the fact at t_3, is such that it is able to make *no* causal difference to events that occur *between* t_2 and t_3.

So understand *dependence* any way you like, then just see which of (1) or (2) is true of it, and then continue reading. Cases of the no-causal-relevance sort (2) are easy to imagine. Suppose that in the closing minutes of a basketball game Team A is behind Team B by one point. Team A then makes a layup, which gives them a one-point lead. Whether this layup is the *game winning shot* depends on facts about happenings after that shot is made. Specifically, if Team B never recovers the lead and so team A wins, then that shot was the game winner. If Team B does recover the lead – whether or not Team A gets it back again, and regardless of which team wins – then that shot was not the game winner. Being the game-winning shot is a sort of post-hoc honorific.

Here is a clear case where facts concerning happenings in the future with respect to some event determine a feature of that event. But it this feature –*being the game winning shot* – is clearly causally inert until the determining events have transpired.

And since these events don't transpire until the final buzzer sounds, a shot's being the game-winner is a causally inert feature of that shot until after the buzzer sounds. For suppose that it were not, that it had some causal potency. Then we would have a convenient method for cutting dead time off the end of basketball games. Any time a team takes the lead with a shot, just determine *at that time* whether the relevant causal power is present — and causal powers are the sorts of things whose presence is detectable, via their effects. And if we detect this causal power, then we have determined that it was the game-winning shot, and we can end the game early. We will know that the team that made that shot will not lose the lead for the rest of the game. Now clearly this this isn't possible. And this is because this feature has no relevant causal powers.

I will call features (or properties, or whatever) of this (2) sort, features determined by facts concerning events in the future of the feature, but which lack causal relevance for anything before those determining events, *game-winning features*. Such features may have consequences for things *after* the determining events, of course. Whether a player is inducted into the hall of fame, or named player-of-the-game, may well depend on whether that player made the game-winning shot *as such*. It might cause me to lose a bet on who would make the game winning shot. But I wouldn't know whether I won or lost until after the final buzzer, of course.

Cases of the causally-relevant (1) sort are easy to imagine, at least in science fiction. When Marty McFly starts playing the guitar in 1955 at the big dance, his arm begins to fade out, thus making it difficult for him to play. The fading out and then back in is having clear causal consequences at the time, most notably with respect to his ability to play his guitar. And the degree of fade-out, and hence his control over his arm, is being determined by an event in the future relative to his guitar playing, namely, his parents getting married and having Marty as one of their children (or not). An event in the future with respect to Marty's guitar playing is retroactively wiping him from existence, one limb at a time, and this is having causal consequences at the time. Let's call features such as this *McFly features.*

Let's return to the case at hand. The first flash occurred at t_1, the second flash will occur at t_3. The challenge was that the extensionalist has no good answer to the question concerning the nature of perceptual experience at t_2+d, the time at which information about events up to t_2 has been processed by the perceptual mechanisms. The response is that we are now considering is that it is possible to say something about experience at t_2+d, because the experience at this time (or more carefully, the extended experiential episode that has so far progressed up to t_2+d) can have features that are logically dependent on things that happen at t_3 or later. We don't need to delay, or revise our assessment of, any features at t_2. So far so good: re-writing would involve embracing the sufficiency of retentionalism as I argued above, and a delay would collapse it to Dainton's delayed retentionalism. So we are being promised a genuine extensionalism that solves the problem without delay.

The question is, is this feature of experience at t_2+d a game-winning feature (2), or is it a McFly feature (1)? The problem is that those options are exhaustive, and neither is viable. McFly features require us to play fast and loose with physical law. I take this to be uncontroversial enough so that I can drop that horn without risking accusations of progressing unfairly. But game-winning features lack causal relevance. And content – phenomenal or subconscious – is the sort of thing that must have causal relevance.

Why should we worry about whether the determined feature has causal relevance? Well, a couple of different considerations come to mind. First, if this feature of experience is not causally relevant, then all motivation for its accuracy immediately goes poof. As does any premium on its timeliness. If it has no causal relevance, why would the perceptual system go though the trouble, rather than just delaying?

Second, it is not at all obvious what sorts of things could count as perceptual content, but lack causal relevance. If there were such things, the following would have to be true: there are a pair of contents – say i) *the screen is black*, and ii) *the screen has a bright dot moving across it* – such that whether I am experiencing (i) or (ii) has no causal relevance. For example, I could not be trained to push a button when I experience (i) but not (ii); (ii) does not elicit any eye-tracking response for the dot I am perceptually experiencing; neither is more likely than, or capable than, the other to elicit a verbal response from me to the effect that I just saw motion. And so forth. Maybe there is something in the vicinity here worth calling a "feature of experience" that I'm just unable to think of.

5.3.2 Ambiguity again

Phillips offers some assistance in understanding what he has in mind:

> … a simple analogy may help. Facts about what I am doing at some instant may depend on what I am doing over some period of time. Thus, whether I am walking or running at some instant is not fixed by a snapshot of my posture at that instant. Nor indeed is it fixed by what I do over a very brief period, e.g. 1/10th second, surrounding that instant. This is clear if we consider the mechanics of walking. As a textbook puts it, 'walking can be characterized as an alternating sequence of single and double support' in contrast to, say, running which 'involves alternating sequences of [single] support and nonsupport' (Enoka, 2002, p. 179). Thus, a single support phase (which is all that will be going on during certain sub-periods of periods of walking) is insufficient an occurrence on its own to determine whether someone is walking over that period. Nonetheless, someone can be walking or running at some moment in virtue of what they do over some extended period of time encompassing that instant. The metaphysically basic units of walking are significantly extended in time. The same, the Extensionalist should insist, is true of experiencing. (Phillips 2011, p. 398)

This example makes the situation sound more promising for the extensionalists than I think it really is. Here is why. It sure seems like the processes of running and walking are very different, even to the point of having different causal powers – running will get me to the bus before it departs but walking won't. What could be more causally legit than that? And it also seems plausible that there can be situations such that whether someone is, at time t_2, running or walking, might be determined by something that happens at t_3. And when you put these two points together, the case looks promising.

But there is an ambiguity in Phillips' example. There are two possibilities: i) the body's *dynamical state* at t_2 – the single support phase – is such that it would be difficult or impossible for the second foot to contact the ground before the other foot lifts. It is important to consider the dynamical state, including the momentum of the parts and torques and tensions on the joints, and not just the instantaneous kinematic description of "1 foot on the ground". An Olympic sprinter at a point in time in the middle of a race might have one foot on the ground, but the large momenta and forces and torques in play at that time all but determine (modulo falling pianos and such) that the planted foot will lift off the ground before the other foot lands. The instantaneous *kinematic* state might be ambiguous, but the instantaneous *dynamic* state is not. And surely when it comes to determining whether someone is walking or running at a time, the dynamic state description should trump the kinematic state description. The dynamic state description gets the right answer when we ask whether the Olympian half way through her race was running or walking when the piano fell on her head. The falling piano did, after all, force her other foot down to the ground while the first foot was still grounded. But her body's dynamical state at that time were geared towards running. Would we really want to say that she was walking when the piano landed on her in the middle of the sprint, just because the result forced her second foot down before her first lifted from the ground?

Consider the "Froude Number", defined as $(mv^2/l)/mg$ and in algebraically reduced form v^2/gl, where v is velocity of the locomoting person, g is gravity, and l is leg length. This is a quantity determinable at a given instant from dynamical features (especially velocity and gravitational acceleration). As Enoka (2002, p. 187) explains: "We cannot walk at Froude numbers greater than 1.0, because that would mean the centripetal force would exceed the gravitational force. Interestingly, many bipeds, including humans and birds, prefer to switch from a walk to a run at a Froude number of ≈.05…" So to take but one example of dynamical state trumping kinematic state, if you have one foot on the ground (kinematic state), but your Froude number is, say, 1.2 (determined by your dynamical state), you cannot be walking. Your dynamical state *at that time* determines that you are, at that moment, running.

The other possibility is ii) the body's dynamical state at t_2 really is ambiguous, meaning that the person could at t_2 start speeding up so that the first foot lifts before the other falls, or slow down a bit so that the second falls while the first is still

on the ground. That is, even if you knew the dynamical state you would not know if that state was part of a running trajectory or a walking trajectory. It is on the cusp and could go either way. Perhaps this is the situation at Froude numbers of a bit less than .05.

In the first case, while it is true that whether the instantaneous state at t_2 is one of walking or running might, in some sense, be determined by what happens at t_3, it is also the case that the relevant event at t_3 (the foot lifting before the other touches) is caused by the dynamical state at t_2. The dynamical state *at* t_2 isn't really indeterminate with respect to running vs. walking at all.

What about the second case, where the locomoting person is in a transitional dynamical state that is on the cusp of evolving into either a run or walk? In this case it is indeed true that whether this state is a phase of a walking vs. running process will be determined by events in the future relative to that state. *But exactly because of this, the 'running' vs 'walking' feature in this case is a game-winning feature.* The instantaneous dynamical state at t_2 has exactly the same causal relevance at t_2 whether it gets determined, by the future event, to be a phase of a walking event or a running event. Indeed, that is exactly what allows it to be ambiguous.

So while the analogy might have initially seemed helpful, once it is unpacked its helpfulness fades. Exactly how a future event can determine an earlier feature without backwards causation that isn't just a game-winning feature is still unclear. And how anything that is a mere game-winning feature of experience can plausibly count as content is also unclear.

6 Discussion

I have tried to accomplish four things. First, I have argued that temporal experience may not be scale invariant, and that because of this, there are hybrid options potentially available that have not been explored in the literature. Second, I have clarified that my own view, the TEM, is meant to apply only at small time scales, with the further specification that it attributes B-ish temporal content to experience at that scale. It is possible that this view could be combined with more standard retentionalist account, or even extensionalist account, of temporal experience at larger time scales, perhaps one that attributes in A-ish contents at these larger scales. I believe that a standard retentionalist picture at larger time scales is probably the way to go, but making that part of the case has been beyond the scope of this paper. And I don't feel like I have a compelling stake in the qustion one way or the other.

Third, I have tried to show that the theory I described, the TEM, is not subject to Dainton's *surplus content* objection. I argued that because the TEM attributes B-ish content to the relevant experience in such a way that no moving now-point is

required to traverse the interval in order to imbue it with temporal content, it is not forced to recognize any objectionable surplus content. What would be required to force the reductio conclusion would be either i) A-ish contents that required a now-point to traverse the entire temporal extent of each trajectory estimate, or ii) a summing of B-ish intervals that double-counts all overlaps. And the TEM isn't forced to do either. It explicitly denies A-ish contents at that scale, and any counting scheme that would require that I claim that I am seeing two meters of oak when looking with both eyes at a one meter oak board is independently objectionable, even if according to some way of summing magnitudes if could be construed as 'correct'.

Finally, I have argued that the three main extensionalist theories face a serious challenge form the problem of temporal illusions at small time scales, namely the scale of around 200 msec, which is where those illusions manifest. Dainton's delayed extensionalism requires all of the machinery of the TEM, but puts that machinery to much less efficient use. In particular, it must claim that all experience is constantly subject to a delay when there would seem to be no benefit in doing so. Hoerl's appeal to "neutral" experiences either collapses to Dainton's delayed extensionalism, or is positing experiences with quite puzzling and seemingly contradictory contents. Philip's reply would seem to be able to account for game-winning features of experience at best, but these don't seem to be the sorts of features one would want from any experiential content worth the name. It is of course possible that despite my best efforts at charitability I have not fully understood Dainton's or Hoerl's or Phillip's views correctly. In that case, my objections should be read as an invitation for clarification and further specification.

Where does all this leave us? My discussion of scale invariance was an attempt to clarify the scale at which I intend the TEM to apply and the character of the temporal content that the TEM attributes to experience at this scale. This restriction of the applicability of the TEM to small temporal scales of less than 200 msec opens the possibility of a hybrid view combining the TEM at small temporal scales with another view, possibly extensionalism or a different flavor of retentionalism, at larger scales. My defense of the TEM against Dainton's objections are relevant to the TEM as applied at small scales. But more to the point, my argument *against* extensionalism – the problem of temporal illusions – applies only to an extensionalism taken to apply at brief temporal scales, since it is only at these scales that temporal illusions are in effect. An extensionalism that was taken to apply only at temporal scales greater than 200 msec or so (but ceded the sub-200 msec scale to some other account) would not be subject to that objection.

ACKNOWLEDGEMENTS

I would like to thank the participants of a 2014 graduate seminar on temporal **perception at UC San Diego for valuable feedback on earlier versions of this paper: Chirag Barai, Claudi Brink, Scott Fernandez, Aaron Finley, Noel Martin, Matthew Piper, Dean Tracy, Brian Tracz and Jason Winning.**

REFERENCES

Dainton, Barry (2008). Sensing change. *Philosophical Issues* 18:362-384.

Dainton, Barry (2013). Temporal Consciousness. *Stanford Encyclopedia of Philosophy*.

Dennett, D. and Kinsbourne, M. (1992). Time and the Observer: The Where and When of Consciousness in the Brain. *Behavioural and Brain Sciences* 15:183–247.

Enoka, R. M. 2002: *Neuromechanics of Human Movement.* Champaign, IL: Human Kinetics.

Grush, R. (2005), Internal models and the construction of time: Generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions. *Journal of Neural Engineering*, 2(3): 209–218.

Grush, Rick (2007). Time and experience. In Thomas Müller (ed.), *The Philosophy of Time*, Frankfurt: Klosterman.

Grush, Rick (2008). Temporal representation and dynamics. *New Ideas in Psychology* 26(2):146-157.

Hoerl, Christoph (2009). Time and tense in perceptual experience. *Philosophers' Imprint* 9(12).

Husserl, Edmund (1991), *On the Phenomenology of the Consciousness of Internal Time* (1893–1917), edited and translated by J.B. Brough. Dordrecht: Kluwer.

Kolers, Paul A., and Michael von Grünau (1976). Shape and color in apparent motion. *Vision Research* 16(4):329-335.

Lee, Geoffrey (2014). Temporal Experience and the Temporal Structure of Experience. *Philosophers' Imprint* 14(3).

Phillips, Ian (2011). Perception and Iconic Memory: What Sperling Doesn't Show. *Mind & Language* 26(4):381-411.

Phillips, Ian (2014). Experience of and in time. *Philosophy Compass* 9(2):131-144.

Rao, Rajesh, David Eagleman and Terrence Sejnowski (2001). Optimal Smoothing in Visual Motion Perception. *Neural Computation* 13:1243-1253.