

# How to, and how *not* to, bridge computational cognitive neuroscience and Husserlian phenomenology of time consciousness

Rick Grush

Received: 6 July 2006 / Accepted: 8 August 2006  
© Springer Science+Business Media B.V. 2006

**Abstract** A number of recent attempts to bridge Husserlian phenomenology of time consciousness and contemporary tools and results from cognitive science or computational neuroscience are described and critiqued. An alternate proposal is outlined that lacks the weaknesses of existing accounts.

**Keywords** Time consciousness · Husserl · Trajectory estimation · Representational momentum · Temporal illusions · Specious present

## 1 Introduction

Over roughly the past 10 years, attempts to build bridges between current computational cognitive neuroscience and Husserlian phenomenology of time consciousness have evolved into an increasingly fashionable endeavor. Described in these general terms the enterprise is a laudable one (I will say a few words at the end of this paper as to why this is so, in case it is not obvious). The problem is that most of the proposed bridges support no actual theoretical weight, confusions and loose metaphors being the materials from which they are constructed. In this paper I will discuss a number of these attempts—ones paradigmatic of the approaches taken—and try to diagnose, as clearly as possible, how and why they misfire. I will then indicate an approach that I think holds promise.

In Sect. 2, I will very briefly discuss those aspects of Husserl's program that will be relevant to the subsequent discussion. Husserl's position is not simple, involving many interacting facets, some of them essentially impervious to comprehension. To make matters worse, his position was continually evolving, and it is not always an

---

R. Grush (✉)  
Department of Philosophy, University of California,  
San Diego, P.O. Box 0119,  
9500 Gilman Drive, La Jolla,  
CA 92093-0119, USA  
e-mail: rick@mind.ucsd.edu

easy matter to discern evolution from inconsistency. These challenges aside, Husserl's analysis is rich, groundbreaking, and an invaluable source of insight. My discussion in Sect. 2 will of necessity be very brief and caricaturish. I'm confident, however, that the sketch will contain sufficient detail and accuracy for the purposes at hand.

In Sect. 3, I will discuss three proposals to bridge various tools from computational cognitive neuroscience and Husserlian phenomenology of time consciousness. In all cases, the core of the proposal is the same—(i) this or that theory or analysis from some area of cognitive science or neuroscience imputes a certain kind of structural feature to patterns of neural activity; (ii) language that can be used to describe these structural features can (sometimes only if aided by an alarming degree of poetic license) also be used to describe aspects of the phenomenology of time consciousness as Husserl analyzes it; therefore (iii) some explanatory or otherwise interesting connection between neural mechanisms and phenomenology has been revealed. And a diagnosis in all three cases is the same: content/vehicle confusion. The material in (i) concerns vehicle properties, never content properties, while the material in (ii) concerns contents, never vehicles, and so one can get a *direct* route from (i) and (ii) to (iii) only via confusing vehicle and content. (Of course, there are indirect routes from vehicle to content, as I will describe in Sect. 4.)

After discussing this core confusion that infects all three proposals, I turn to a more detailed discussion of each, since each has additional features and problems. The first example is from Timothy van Gelder (1996), who sees in dynamical system theoretic approaches to cognitive phenomena reflections of aspects of Husserl's program. The second example is from Francisco Varela (1999), who appeals to some specific temporal properties of dynamic coupled oscillators. Finally, and most recently, Dan Lloyd (2002) has attempted to discern, via some sophisticated and independently interesting methods for mathematical analyses of fMRI data, patterns in neural activation that might correspond to elements of Husserl's scheme.

In Sect. 4, I will turn to the issue of what would be required to do an adequate job of discerning the neural substructure of Husserlian phenomenology. Specifically, what is needed is a middle-level theoretical framework that can serve to genuinely bridge, without reliance on metaphor, both (i) the temporal profiles of *content* structures, and (ii) instantiating physical machinery, or *vehicle* properties. I sketch a proposal for exactly such a framework.

In Sect. 5, I conclude with discussion of two issues. The first is some respects in which the proposal I outlined in Sect. 4 fail to fully match elements of Husserl's program. The second issue is a brief identification of an important theoretical stance that is underappreciated by the vast majority of those who work on understanding the nature of the physical bases of consciousness and cognition, a stance on which I am an ally of those I criticize in the earlier sections of the paper.

## 2 Relevant aspects of Husserl's phenomenology of time consciousness

Husserl's *Lectures on the Consciousness of Internal Time*<sup>1</sup> presents many challenges. Husserl's characteristic opacity is, in the case of *this* text, layered atop the fact that the text itself was never prepared by Husserl as coherent book. The text is culled

<sup>1</sup> When discussing Husserl's doctrines I will refer exclusively to *The Lectures on the Consciousness of Internal Time*, from the volume *On the phenomenology of the consciousness of internal time (1893–1917)*, a translation by John Brough, of *Husserliana Band X* (Rudolph Boehm, ed.).

from notes that Husserl wrote on the topic of time consciousness over a period of at least 16 years, from 1901 to 1917. A text extracted from copious notes would be bad enough, but the extraction was not even done by Husserl himself, but rather by his long-time secretary Edith Stein and the editor of the *Lectures*, Martin Heidegger. Furthermore, Husserl's position evolved greatly over this period, this evolution not explicitly marked in the resulting text at all, with the result that the doctrine can appear to be confused and even self-contradictory. For purposes of this paper, however, we can avoid all of the subtleties introduced by these factors, and focus on a few relatively well-defined aspects of the doctrine: the tripartite structure of time consciousness as consisting of retention, protention, and primary impression; the recursive nature of successive now-consciousnesses; and the 'absolute time-constituting flow'.

Two preliminaries before introducing these doctrines. First, Husserl begins with his famous *phenomenological reduction* in which he announces that the topic of his investigation is not actual physical objects, but is restricted to appearances. By this Husserl means roughly that he wants to focus exclusively on the *contents* of conscious awareness, and not worry about any real objects which may or may not correspond to such appearances. So the expression 'object' is taken to mean an object *as something conceived by or presented to/in* a conscious mind. Convincingly hallucinated objects are thus objects in the relevant sense. Second, Husserl makes clear that he is interested in temporal objects. These are objects of consciousness that are given *as being in*, or *enduring through time*. A favorite example is a melody, but even rocks and trees are temporal in that they are experienced as persisting through time. These would contrast with things like abstract objects, such as numbers or the Pythagorean theorem. Though even in such cases temporality is not entirely absent, since a conscious agent can be aware of the fact that its own contemplation of the Pythagorean theorem is something that takes place in time. We won't, however, be concerned with the more sophisticated elements of Husserl's analysis that deal with a subject's experience of itself as temporal.

With these preliminaries in hand, we can turn now to the relevant doctrines, and first to the tripartite structure of time consciousness (Husserl's discussion of this is largely in Sects. 8–13 of the *Lectures*). A common conception, one Husserl will denounce, of the content entertained by a conscious mind is that it consists of a series (if discrete) or stream (if continuous) of conscious contents that mirror to some degree of accuracy events in the environment. In particular, this mirroring is assumed to be isochronic in that at each objective point in time, the content that is entertained by the mind corresponds to the state of the experienced situation at that time, perhaps with a slight delay introduced by neural or psychological processing. So for example, if you are watching a bowling ball strike the pins at the end of the lane, at the instant the ball impacts the first pin, the relevant content of your perceptual episode is something like *the ball in contact with the first pin*, and nothing about where the ball has just been, or what it is about to do, is part of this perceptual content. A proponent of this view need not claim that the ball's previous motion and its imminent motion are completely outside the mind's reach. The mind has *memory* and can formulate *expectations*. The claim is simply that none of this is part of the *perceptual* content at that instant.

On Husserl's analysis, however, perceptual content is not temporally punctate in this way. That aspect of your perceptual content that corresponds to what is happening at that instant—the ball's being in contact with the first pin—is just that: *one aspect* of your perceptual experience. Two other aspects, dubbed *retention* and *protention*, concern what you have just experienced and what you, in a specific sense

to be discussed shortly, *expect* to experience. Husserl's tripartite structure of time-consciousness is accordingly composed of these three aspects—*primal impression*, which corresponds to what is new at the strictly present instant; *retention*, which corresponds to the content from recent experience that is retained in consciousness and provides, among other things, a past-directed temporal context for primal impression; and *protention*, which corresponds expectations of the imminent course of experience which provides, among other things, a future-oriented temporal context for primal impression.

Husserl is keen to insist that retention and protention are *perceptual* in nature, and for what follows it will be useful to explore what he means by this and his reasons for insisting on it. It will be most convenient to focus on retention, since the case for protention is less clear and less worked out, but to the extent that it is, it is clearly supposed to be essentially parallel to retention. Consider a melody, such as the main theme from the 4th movement of Beethoven's *Ninth Symphony*, and in particular the first five notes: C<sup>#</sup>, C<sup>#</sup>, D, E, E. These five notes are exactly repeated in the 3rd bar of the theme. Now when these five notes recur in the 3rd bar, there is a sense in which you are experiencing the same thing you experienced two bars back, namely the note sequence C<sup>#</sup>, C<sup>#</sup>, D, E, E. This part of the series of primal impressions is the same. Nevertheless, the experience of these notes is different. The reason for the difference is the temporal context. At the time the third bar begins, the first and second bars have already played, and they thus set a temporal context for the third bar that was not present at the time the first bar played. Especially to anyone familiar with the symphony, the beginning of the third bar sounds distinctively different from the beginning of the first bar, despite the fact that they strictly consist of an identical sequence notes. The difference is a difference in the perceptual content grasped in the two cases, and is accounted for by Husserl by the fact that perceptual content is not exhausted by primal impression, but includes retention. 'Retention' and 'protention' are Husserl's names for the processes that provide this perceptual temporal context. When the 3rd bar begins, the first two bars are not entirely wiped from consciousness. The first five notes of the 1st bar are heard *as initiating* the melody. The first five notes of the 3rd bar are *heard as* occupying a different location in the melody, and hence as doing different work in the melody.

This element of Husserl's program as I have just described it will probably not invite much resistance, since the idea that the mind has memory is hardly a matter of controversy. However, Husserl is insistent that retention is unlike memory as it is typically conceived (the discussion of this issue is primarily in Sects. 14–24 of the *Lectures*). Husserl's proposal is not that in the normal case of listening to the symphony, at the time the third bar begins playing, one *recollects* the first bar. Such *recollection* would be one way to understand memory, as a *re-experiencing of some past experience*. One *can* engage in this exercise of recollection, of course. When the 3rd bar begins you can recollect the 1st bar, and create a complex experiential state that consists of an amalgam of your present perceptual experience together with a recollection of something you recently heard. Such an amalgam would be similar in some respects to hearing two symphonies at the same time, one playing the first bar and the second playing the third bar. But this isn't the way in which, in the normal case, the 1st bar sets up in consciousness the context in which the 3rd bar is heard. In the normal case, the first bar, after it sounds, is not drudged up again to be re-experienced. Rather (there are, unfortunately, extremely limited linguistic resources for describing this

phenomenon) it sinks back from the ‘now’, while remaining in consciousness, to form something analogous to a visual periphery.

Furthermore, if I *do* recollect the 1st bar, then my recollection will, if I endeavor to make the imagery convincing enough, exploit the sort of temporal structure of consciousness that it is Husserl’s task to explain. When I *recollect* the first bar, the recollection has a temporal course, it will take approximately the same amount of time to complete the episode of recollection as the original experience took. And as the recollection progresses the series of recollected notes proceeds in a sequence such that at the time I recollect the second note, the first note, which was just recollected a moment ago, has left its mark on retention, and thus provides a context for the currently recollected second note. Memory, understood as recollection, is a process that *exploits*, rather than *explains*, the tripartite structure of experienced temporality. This is what Husserl means by classifying retention as *perceptual*. Retained contents are an aspect of our perceptual experience, not a matter of memory understood as recollection.

In addition to retention and primal impression, we have protention which concerns imminent experience. If we recognize the five note melody, then at the time the fourth notes sounds we not only hear the fourth note (primal impression) in the context of the just-past notes (retention), but we have an expectation of the about-to-be-heard note—protention. But protention is not limited to this sort of case. When you are looking at a tree, you have specific protentions regarding the tree, in this case the relatively boring expectation of its continuing existence. If it suddenly vanished that would be rather surprising, the surprise being a violation of your protentions of continued tree-experience. Though protention is announced as one of the three integral aspects of time-consciousness, it is relatively neglected in the *Lectures* themselves. And for the main purposes of this paper, we will be able to set protention aside. Primal impression and retention will serve. I will however return to the topic of protention in the final section of this paper.

The next Husserlian doctrine is what might be called the recursive nature of present-time consciousness (this is discussed mainly in Sects. 27–29 of the *Lectures*). The process just described above is one in which what was the content of primal impression becomes the content of retention. This is, according to Husserl, something of a simplification. A better characterization would be that at each moment the entire content of conscious awareness sinks back via retention. So the content of my retention of what happened a brief moment ago is not just what was primal impression a moment ago, but the full retention–impression–protention structure from that moment. The retentional phase of each new ‘now’ includes a retention of the previous ‘now’, not just of the previous primal impression. The result is a sort of recursive nesting of nows feeding into nows. My brief gloss on this aspect of Husserl’s doctrine does not do it justice.

The third and final doctrine that will concern us is what Husserl calls the ‘absolute time-constituting flow’ (this is discussed primarily in Sects. 34–40 of the *Lectures*). This is motivated by the following issue. The description of protention, retention, and primal impression, and their relation, described them in temporal terms—primal impression *becomes* retention, for example. This suggests that these structures of time consciousness are themselves experienced as being within some distinct temporal flux. Husserl maintains that this suggestion is inaccurate, however. While the language used to describe the relations between protention, retention and primal impression impute temporal characteristics to them, this is due to expressive

limitations of natural language. These relations are not relations among items that are located within an independent subjective temporal flow—rather, these relations *constitute* the flow of subjective time. Husserl makes the point by distinguishing (i) the objects that *are constituted* as temporal objects by the way that they are structured by protention, retention and primal impression, from (ii) the relations between various ‘phases’ of consciousness that *constitute* the temporality of temporal objects.

Time-constituting phenomena, therefore, are . . . fundamentally different from those constituted in time. . . . it . . . can make no sense to say of them (and to say with the same signification) that they exist in the now and did exist previously, that they succeed one another in time or are simultaneous with one another, and so on. But no doubt we can and must say: A certain continuity of appearance—that is, a continuity that is a phase of the time-constituting flow—*belongs* to a now, namely, to the now that it *constitutes*; and to a before, namely, as that which is constitutive (we cannot say “was”) of the before. But is not the flow a succession, does it not have a now, an actually present phase, and a continuity of pasts of which I am now conscious in retentions? We can say nothing other than the following: This flow is something we speak of *in conformity with what is constituted*, but it is not “something in objective time.” It . . . has the . . . properties of something to be designated *metaphorically* as “flow”; of something that originates in a point of actuality, in a primal source-point, “the now,” and so on. . . . For all of this, we lack names. (Husserl, 1991, p. 79)

With these brief remarks in hand, we can turn now to some recent attempts to bridge Husserlian doctrine and contemporary cognitive science and neuroscience.

### 3 Three recent proposals

#### 3.1 Introductory

As I mentioned briefly in the introduction, the three proposals that I will be discussing are all guilty, among other things, of content/vehicle confusions. It will be helpful to say a bit about what contents, vehicles, and confusions between them, are.

The 19th Century visual physiologist Ewald Hering dubbed the dark grey that one appears to be visually presented with in the absence of light ‘brain grey’, a sort of neutral resting point of the opponent processes that normally produce specific color and brightness experiences via contrasts (for interesting discussion, see the first chapter of Clark, 2000; and Sorensen, 2004). It turns out that the cortex, including the visual cortices, are grey. Now it would be an obvious blunder for anyone to claim to have said anything even remotely interesting about brain grey, let alone to have explained it, by pointing out that the neural hardware of the visual system is grey. To do so would be to commit a blatant content/vehicle confusion, a confusion characterized by the attempt to read features of the *content* carried by a representation directly from analogous features of the *vehicle*—the material substrate—of the representation. It may not be quite as obvious, but content/vehicle confusions are still confusions even when the topic is temporal content. The word ‘Tuesday’ means Tuesday, even if it is written on Monday. And if the word ‘Tuesday’ is written on Tuesday, the fact that it is written on Tuesday has nothing to do with why it means Tuesday.

As a corollary of the fact that properties (to put it loosely) of contents cannot be read directly off the properties of the supporting vehicles, similarity relations

between multiple contents cannot be read off any similarity relations between the properties of the supporting vehicles. The inscription ‘ten’ shares many features with the inscription ‘tan’, many more than it shares with the inscription ‘eleven’, at least on any non-pathological measure of similarity. It would obviously be rash to conclude from this that the meaning of the word ‘ten’ means something very close to what ‘tan’ means, closer anyway than to what ‘eleven’ means. The issue is more stark if one considers the vehicle and content similarities between the inscription ‘ten’ and the inscription ‘the square root of sixteen, factorial, divided by two, minus the square root of four’.

And it is not only linguistic representations for which these facts are true. Consider a binary representation of numbers implemented as the presence or absence of charge in a set of 11 capacitors, and consider the binary representation of the numbers that would be represented in base ten as ‘1’ and ‘1024’: namely ‘0000000001’, and ‘1000000000’. A natural and common measure of the difference between these binary representational *vehicles* is the *Hamming distance*, which is just the number of *different* binary digits—in this case, the number of capacitors that have a different charge state. Here the Hamming distance is 2, since only two capacitors (the first and last) have a different charge state in the two representations. Nine of the 11 capacitors have the same charge. Consider next the Hamming distance between the binary representations of 1024 and 1023: ‘1000000000’ and ‘0111111111’. The difference in the vehicles is maximal—a Hamming distance of 11. They differ at every spot where a difference in the vehicle could matter, every one of the capacitors has a different charge state. Now suppose one were to measure the physical properties of the set of capacitors, and use this measure as an index of what was being represented. One would conclude that ‘1000000000’ carried a very different content from ‘0111111111’, and carried a very similar content to ‘0000000001’, an obviously bad conclusion.

These brief remarks on content/vehicle confusions should suffice for now. I will return to content/vehicle issues in Sect. 3.5, where I will discuss how one might go about trying to save the inference, in at least some cases, from vehicle properties and relations to corresponding content properties and relations.

### 3.2 van Gelder’s ‘dynamical systems theoretic’ proposal

Timothy van Gelder (1996) has argued that cognitive science and phenomenology can mutually inform each other, and uses as a test case a specific dynamical systems theoretic model of auditory pattern recognition, the *Lexin* model (Andersen, 1994) and Husserlian analyses of time consciousness. In this section I will first discuss the Lexin model, and then discuss the comparisons van Gelder makes between this model and Husserlian phenomenology.

There is a minor exegetical challenge here in that the actual Lexin model has features that are quite unlike those van Gelder attributes to it. To avoid being hindered by this issue, we can be maximally charitable to van Gelder by defining an alternate model, the Lexin\* model, to be a model that fits van Gelder’s description. Doing so will let us test van Gelder’s proposal as he intends it, rather than getting bogged down by the less important fact that van Gelder’s actual inspiration does not fit his description of it.<sup>2</sup>

<sup>2</sup> Briefly, van Gelder’s description the Lexin model takes it to be a standard dynamical system that differs from classical system in that it does not employ memory registers. The actual Lexin model in

In essence, the Lexin\* model is a dynamical system consisting of a set of artificial connectionist units. Each of these units has an activation at each time that can be represented as a scalar value, and the system's state at any time will be the set of all these values. Equivalently, one can describe the system's state at any time as a point in the system's *state space*. The state space will be a 'space' defined by letting each dimension correspond to the range of possible values of each unit. The set of actual values of each unit would then specify a *point* in this space. The changes in the activation values of the units over time can thus be represented as the movement of this point through the state space over time, its *trajectory through state space*. At each time, the model's location in state space is a function of three factors: its location at the previous time, its own tendency to meander certain paths through its state space, and some external inputs. The model's evolution over time can be picturesquely described as a combination of its own intrinsic tendency to trace out specific trajectories through the state space, together with external nudges that push the trajectory in directions in which it may have not gone without the nudge.

The system recognizes auditory patterns as follows: from an initial state, the model gets coded inputs corresponding to the initial stages of some auditory stimulus, such as the opening notes of a melody. The model *learns* to be such that a specific sequence of inputs pushes it into a unique region of its state space. The model's location in such a proprietary region of its state space, either during or at the conclusion of the auditory pattern, constitutes its recognition of that specific pattern. As van Gelder puts it

In the Lexin model, particular sounds turn the system in the direction of unique locations in the state space . . . When exposed to a sequence of distinct sounds . . . the system will head first to one location, then head off to another, waving and bending in a way that is shaped by the sound pattern, much as a flag is shaped by gusts of wind. . . (van Gelder, 1996, §23)

Now, how can an arrangement of this kind be understood as recognizing patterns? Recognizing a class of patterns requires discriminating those patterns from others not in the class, and this must somehow be manifested in the behavior of the system. One way this can be done is by having the system arrive in a certain state when a familiar pattern is presented, and not otherwise. That is, there is a "recognition region" in the state space which the system passes through if and only if the pattern is one that it recognizes. Put another way, there is a region of the state space that the system can only reach if a familiar auditory pattern (sounds and timing) has influenced its behavior. Familiar patterns are thus like keys which combine with the system lock to open the door of recognition. (van Gelder, 1996, §24)

The next task is to see what parallels can be discerned between this model and Husserl's analysis of time consciousness, retention in particular. The master thought is that Husserlian retention is that by which past experiences are retained in current

---

Footnote 2 continued

fact employs what are effectively memory registers. As Sven Andersen (creator of the Lexin model) puts it in his dissertation, the "LEXIN model network uses . . . stimulation from time-delayed lateral connections to acquire salient acoustic transitions in the environment" (pp. 47–48). Earlier on, this memory is described as follows: "A delay line spanning some duration that is tapped at particular points acts to create a spatial array from a temporal sequence . . ." (p. 36). This 'spatial array' is effectively a memory register.

consciousness, and the Lexin\* model is touted as translating this into solid scientific terminology.

How is the past built in? By virtue of the fact that the current position of the system is the culmination of a trajectory which is determined by the particular auditory pattern (type) as it was presented up to that point. In other words, retention is a geometric property of dynamical systems: the particular *location* the system occupies in the space of possible states when in the process of recognizing the temporal object. It is that location, in its difference with other locations, which “stores” in the system the exact way in which the auditory pattern unfolded in the past. It is how the system “remembers” where it came from. (van Gelder, 1996, §38)

So the fact that the model could only be in its current location in virtue of having traced out some specific trajectory in the immediate past is the basis upon which the past is being described as ‘an aspect’ of the current state of the system. This is something of a slide to be sure, and not just because of the obvious point that if this were a good analysis every physical system in the universe would exhibit Husserlian retention. Husserlian retentions are aspects of the current contents of awareness in the sense that they are presently existing states that are occurring in consciousness along with states corresponding to primal impression and protention.<sup>3</sup> In the Lexin\* model, the past states of the system are by no means ‘aspects’ of the system’s current state in this sense any more than the cup’s being on the edge of the table is an ‘aspect’ of its subsequent location on the floor. From the standpoint of Husserl scholarship and interpretation this slide is *colossal*. Husserl’s program is motivated, in large part, by the realization that a mere sequence of conscious states is not sufficient for consciousness of a sequence; that, e.g., hearing a melody as a melody cannot be reduced to a mere sequence of note hearings. Counting what was *in fact* a prior state as an ‘aspect’ of the current state, and identifying this as ‘retention’ essentially jettisons everything of interest in Husserl’s program, so far as I can tell, even if that jettisoning is cloaked in the slop-expression ‘aspect’.

The second (and to some extent the fifth) point of comparison van Gelder draws between the Lexin\* model and Husserl’s analyses is degree of pastness. The proposal is to exploit the idea that the system is causally sensitive to the order of inputs, in that providing A and then B as inputs will push the model into a different region of state space than providing B and then A as inputs. And the fact that there are temporal limits to retention is tied to the idea that over time the effect of influences on the state of the system get ‘washed out’.

In these kinds of dynamical auditory pattern recognition systems, the location of the current state of the system reflects not just *what* previous sounds it had been exposed to, but also the *order* in which it was exposed to those sounds – or, more generally, *how long ago* it was exposed to that sound. Therefore, there is a clear sense in which retention, on this interpretation, intends the past as past to some degree. (van Gelder, 1996, §40)

It is a fact about these dynamical models that influence is “washed out” in the long run. Generally, the longer the state of the system is buffeted about by

<sup>3</sup> I use the expression ‘state’ for convenience. Husserl prefers ‘phase’, since ‘state’ suggests that they are independently manifestable. Husserl is keen to insist that retentions and protentions are in fact phases of a continuum that can only manifest as aspects of the whole continuum. The point is that whether you call them states or phases, they are entities that co-manifest.

current inputs, the less the influence of a past input is discernible in the current state. There are not strict limits here, but it is clear that the current state does not retain the influence of inputs arbitrarily far in the past; thus, the current interpretation confirms the phenomenological insight that retention is *finite*. (van Gelder, 1996, §40)

Any sense of adequacy here is maintained only by either not recognizing poetic license for what it is, or by embracing an alarming content/vehicle confusion. The content/vehicle confusion embodied in the inference from fact that the vehicle has a temporally prior causal precursor sequence to the conclusion that any content carried by the vehicle is content to the effect that there was such a sequence, should be clear enough that further comment is not needed. As to the finitism of retention being cashed out in terms of causal efficacy: note that the gust of wind may have exerted its influence on my car's trajectory after the speeding freight train did, but the temporal order of the influences clearly cannot be read off the strength of their influence. The same is true of the Lexin\* model. Early notes may well have constrained subsequent trajectory significantly more than later notes. In order for this proposal to be adequate, the degree of influence that an input has on a given state space location would have to be a strictly monotonic (decreasing) function of the temporal interval between the time of the influence and the time of the state in question. This condition obviously fails to hold, even in the Lexin\* model.<sup>4</sup>

The sixth point is van Gelder's claim that retention is 'direct', and this gets glossed as "not memories, images, or echoes." This gets spelled out more fully

How can I be conscious now of something which is not now and hence, in a sense, does not actually exist? One possibility is that I am conscious of another thing which does exist now, and which has the function of (re)presenting that temporal stage. This kind of consciousness of the not-now is indirect; it travels via something that is now. Husserl, however, is adamant that this is not how retention intends past temporal stages. Retention is not a matter of having images of a past stage. Retention is not having memories which recreate the past as if it were now; nor is it like echoes still hanging around in the now. As Brough puts it: "Retention does not transmute what is absent into something present; it presents the absent in its absence" (276). Retention reaches out directly into the past. It is more like perceiving than representing. (van Gelder, 1996, §14, van Gelder's reference is to Brough, 1989, p. 276)

I will comment on this passage shortly. First, the connection between this and the Lexin\* model

Clearly, retention as explicated here does not relate to past stages of the temporal object via some other current mental act, such as a memory or an image or an echo. There is no space for any such additional acts in the model. In that sense, retention is *direct*. (van Gelder, 1996, §40)

Here van Gelder is executing maneuvers aimed at putting the shortcoming discussed under item one in a positive light. The criticism I produced in discussing the claim that retentions were *present* was that there was nothing in the Lexin\* model, no genuine *aspect* of its state at a time, that was *representing* any of its past states. But if

<sup>4</sup> Van Gelder does say that "It is a fact about these dynamical models that influence is "washed out" in the long run" (§40), but I can find nothing in the Lexin or Lexin\* models to explain why he thinks this.

van Gelder's gloss here on Husserlian retention to the effect that it is 'direct' and not mediated by another present state, is correct, then the fact that the Lexin\* model implements retention only in the causal-precursor sense of 'aspect' might be a good thing.

The problem with this line of thought is that the characterization of Husserlian retention on which it relies is importantly inaccurate. Retention for Husserl is indeed distinguished from memory proper, as I discussed in Sect. 2. But the difference between retention and memory is not that memories are current content-bearing states and retentions are not, for both are current content-bearing states. Rather the difference is in how they present their content. A memory re-presents the content in the sense that it reconstructs a sequence of experiences and exhibits them again, yielding another temporal experience with its own tripartite structure. By contrast, retention is a phase of the current contents of temporal awareness that presents the content it carries as just-past to some degree. Van Gelder's gloss begins with the true characterization that "[r]etention is not having memories which recreate the past as if it were now." But what makes this gloss true, what makes retention unlike memory, is that retention does not re-present its content 'as if it were now'—that is, retention does not, as recollection does, construct at the current time a new, re-presented experience. Van Gelder, however, suggests that what makes the gloss true, what makes retention different from memory, is the fact that memories are current content bearing states and retentions are not.<sup>5</sup> But that is simply inaccurate. In *that* respect, retention is like memory. To put the point another way: Husserlian retention is itself 'indirect' in exactly the sense that van Gelder is here chiding. Retention is consciousness of the not-now that is mediated by something that is now, *a retention*.

In summary, the idea that the Lexin\* model (or any similar dynamical model analyzed in a similar manner) can serve as a bridge to Husserlian phenomenology suffers from a number of fatal shortcomings. First, the proposal conflates the crucial difference between representing something that is past (in the sense of being intentionally directed at it, not in the sense of re-presenting it), and being causally influenced by or even determined by something that is past. When the issue is the *contents* of experience—Husserl's topic—clarity on this difference could not be more paramount. A corollary of this conflation is the requirement that degree to which something is 'intended as past' is a function of degree of causal determination—with causal factors in the past being more and more 'washed out'. Second, Husserl is mis-interpreted at a number of points, and not benignly so. E.g., the suggestion that there is no present state that mediates the relation between consciousness as it is now and something that is presented as just-past *just is* the suggestion that there is no such thing as Husserlian retention.

### 3.3 Coupled neural oscillators

The next proposal is, for lack of a better name, Francisco Varela's coupled oscillator model (Varela, 1999). Varela states the purpose of his proposal thus

<sup>5</sup> Oddly, van Gelder's first two points, that retentions are current, and that they are intentional (read: they carry a content, they are about something) correctly entail that retentions are representations, the denial of his point six. The word 'representation' may be part of the problem here. The relevant meaning of 'representation' for current purposes is something that carries a content, or is about something. However, the word suggests that this job is done by 're-presenting' something in the sense of trotting it out again. But from the fact that retention does not trot anything out again, it should not be concluded that it does not carry a content. This is another way to get a grip on van Gelder's slide.

My purpose in this article is to propose an explicitly naturalized account of the experience of present nowness on the basis of two complementary sources: phenomenological analysis and cognitive neuroscience. (Varela, 1999, p. 111)

The phenomenological analysis is Husserl's tripartite analysis of present time-consciousness and the absolute time-constituting flow. The cognitive neuroscience end of the proposal centers on the dynamics of coupled neural oscillators. A neural oscillator is a small neural pool, perhaps even a single neuron, that cycles through a sequence of states. In the simplest case this might be switching back and forth between two states. *Coupled* neural oscillators are sets of oscillators that are interconnected in such a way that the state of each is influenced by the states of the others. Depending on various factors, it is possible for interesting behavior to emerge from such sets of coupled oscillators, including transient phase locking. An everyday example would be two people carrying a stretcher. Each person's gait is an oscillation—a cycling between states of weight on the left foot and the right foot. When two people are both carrying a stretcher they are coupled in that the movement of one has an effect on the movement of the other through forces mediated by the stretcher that each is connected to. The coupling in this case is such as to produce forces that oppose or enhance the motion of each gait with the result that an in-phase gait is 'rewarded'. In effect this means that even if the two people have different gaits, there will be points at which they will phase-lock for a period of time, the phase locking being the result of the natural tendency of the oscillations to follow a different time course being temporarily overcome by the countering forces produced when the gaits start to decouple. Depending on various features of the system, the oscillators may phase lock permanently, they may lock in counter-phase, or they may cycle through periods of phase locking separated by periods of un-locked oscillation.

A lot of work has been done studying patterns of coupled neural oscillators in the nervous system. The work describes how systems of coupled neural oscillators behave in general, and also how inputs from outside the system (typically sensory inputs, analogous to a third party pushing or pulling on the stretcher as it is being carried) can affect what patterns of transient phase-locking occur.

With this background in hand, we can return to Varela, who introduces three time scales that he takes to be *cognitively* important

At this point it is important to introduce *three scales of duration* to understand the temporal horizon as just introduced:

- (1) basic or elementary events (the '1/10' scale);
- (2) relaxation time for large-scale integration (the '1' scale);
- (3) descriptive-narrative assessments (the '10' scale).

This recursive structuring of temporal scales composes a unified whole, and it only makes sense in relation to object-events. (Varela, 1999, p. 116)

Varela provides some examples of phenomena at these different scales. For the 1/10 range

These elementary events can be grounded in the intrinsic cellular rhythms of neuronal discharges, and in the temporal summation capacities of synaptic integration. These events fall within a range of 10 milliseconds (e.g. the rhythms of bursting interneurons) to 100 msec (e.g. the duration of an EPSP/IPSP sequence in a cortical pyramidal neuron). These values are the basis for the 1/10 scale. Behaviourally these elementary events give rise to micro-cognitive phenomena

variously studied as perceptual moments, central oscillations, iconic memory, excitability cycles and subjective time quanta. For instance, under minimum stationary conditions, reaction time or oculo-motor behaviour displays a multimodal distribution with a 30–40 msec distance between peaks; in average daylight, apparent motion (or ‘psi-phenomenon’) requires 100 msec. (Varela, 1999, p. 117)

Other than a collection of processes that take between about 10 ms and 100 ms it is not clear what this scale is supposed to capture. As far as I can tell, the idea is that basic perceptual discriminations (noticing movement, reaction time) and basic neural events (cell firings) are supposed to operate at this scale. But we turn next to the 1 s scale

A long-standing tradition in neuroscience looks at the brain basis of cognitive acts (perception–action, memory, motivation and the like) in terms of *cell assemblies* or, synonymously, *neuronal ensembles*. A cell assembly (CA) is a distributed subset of neurons with strong reciprocal connections. (Varela, 1999, p. 117)

At this time scale, the psychological and behavioral phenomena are supposed to be things like coherent perceptuo-behavioral events, like uttering a complete sentence or pouring a cup of coffee. Varela also claims that typically the time course over which periods of phase-locking of neural assemblies emerge and dissipate is at this same 1-s scale. And with this

I am now ready to advance the last key idea I need to complete this part of my analysis: *The integration–relaxation processes at the 1 scale are strict correlates of present-time consciousness*. (Varela, 1999, p. 119)

In effect, the fact that an assembly of coupled oscillators attains a transient synchrony and that it takes a certain time for doing so is the explicit correlate of the origin of nowness. As the models and the data show, the synchronization is dynamically *unstable* and thus will constantly and successively give rise to new assemblies. (Varela, 1999, p. 124)

To summarize, analytical, computational and physiological sources suggest that there are patterns among coupled neural oscillators that manifest on the order of about a second, namely transient and inherently unstable phase locking. These patterns arise spontaneously, last about a second, and then spontaneously dissipate, only to be replaced by a different but qualitatively similar pattern. The hypothesis is that these neural events underwrite psychological phenomena at the 1-s scale, and these are the contents of present-time consciousness.

At this point one might well scratch one’s head. On the physical implementation side we have a kind of process—the generation and dissipation of phase-locking in a set of neural oscillators. On the psychological side we have something called present-time consciousness or psychological ‘nowness’. It is natural to suppose that Varela has Husserl’s analysis of time-consciousness in mind here (Varela says as much, and gives an extended discussion of Husserl), but then it is unclear what the relationship is supposed to be, even if we pretend temporarily that slides between content and vehicle are not objectionable. There is nothing in Husserl’s analysis that isolates events at the scale of one second as having any special status. If he had—which he didn’t—then

presumably there are two ways this might have gone. First, perhaps primal impression would manifest on the 1-s scale, and protention and retention would reach out beyond this scale. If this is the proposal, then perhaps Varela means that the prior neural activity before phase-locking corresponds to retention, and future uncorrelated activity after phase-locking corresponds to protention. It is unclear, however, what is protentional about future decoupled oscillators, and what is retentional about previous decoupled oscillators, other than the fact that such states succeed and precede the current state. But this move would be to embrace a content/vehicle confusion. Furthermore, Husserl in fact continually describes time consciousness as characterized by continua, but discrete bouts of phase-locking are anything but continuous.

The other possible connection between the 1-s scale and Husserl's analysis is that the entire tripartite structure of retention, primal impression and protention might be taken to correspond to the 1-s scale. This seems consistent with Varela's suggestion that the *nowness* we are trying to explain corresponds to something like a *specious present*.<sup>6</sup> But then it is still unclear what the relation is supposed to be. Husserl did not think that our time consciousness came in chunks, that we have one conscious episode that includes retention, primal impression, and protention, and then this episode dissolves and is replaced, after about a second, with a new discrete episode of protention/retention/impression time-consciousness. Indeed, this is inconsistent with Husserl's analysis, since the contents of retention at any moment are a continuously transformed version of the contents of primal impression at the previous moment. On the interpretation under consideration here, at each 1-s bloc, a new protention–impression–retention structure is produced anew, and there is no obvious relation, let alone an obvious continuous one, between what was primal impression in one bloc and what is retention in the next bloc.

Note that all of the above puzzles have nothing to do with the deeper difficulty, the implicit content/vehicle confusion that would remain even if we knew what the features of the content and the vehicle were that were supposed to be related. I turn now to the content/vehicle confusion itself. Varela uses bi-stable image—an ambiguous visual stimulus that can be seen in one of two ways—as example of, apparently, experience that has a temporal element. I presume the temporal feature of interest is that there is a time course to the switching between seeing the image one way and seeing it the other way. And indeed there has been empirical work that has claimed to find correlations between, on the side of experience, seeing such an image one way and seeing it another way, and on the neural side, the creation of different patterns of synchrony in neural pools.

But there is no reason to think that this transient phase-locking explains any more than the timing of the process. Worse, there is reason to believe that it *could not* explain any more than the timing of the process. The proposal, recall, is that our time consciousness, our awareness of the flow of time, is to be explained by these patterns of transient synchrony. If the proposal were correct, then the bout of transient synchrony that corresponded to seeing the figure in one way *just would be the subject's temporal unit*. And if this were the case, then by definition the subject should not be able to discern any temporal difference between what might in fact be longer and shorter episodes of image seeing. That is, suppose that I look at the image, and a

<sup>6</sup> The expression 'specious present' was given currency by William James in his *Principles of Psychology* 1890, and was, roughly, the idea that the contents of consciousness any moment spanned a temporal interval. This could naturally be taken to mean that at each moment, consciousness includes a complete protention/retention/primal impression structure.

transient synchrony develops that is a correlate of my seeing the image in one way, and that this neural/perceptual episode lasts 1.5s; and then a new pattern of synchrony develops that is the correlate of my seeing the image in the other way, and this episode lasts .9s. According to Varela's proposal, I should be completely unable to detect this temporal difference. Each episode constitutes, by definition, one subjective temporal now-unit. The capacity to detect a temporal difference would presumably be a matter of assessing the temporal durations occupied by the perceptual episodes and noticing a difference, and this can be done only if the durations of these episodes are assessable against some independent measure. Now clearly you can notice exactly such differences. Look at a Neckar Cube, and get it to switch back and forth, and see if you can discern differences in the amount of time that the cube appeared one way and another. If you can, then you have some reason to doubt that your subjective sense of time is to be explained by the transient neural synchrony phases that co-occur with your image-interpretations in the way Varela suggests.

And so far as I can tell, there is no easy escape from this problem for Varela. He explicitly posits that the self-driven process of emerging and dissipating transient synchronies is the neural correlate of Husserl's absolute time-constituting flow, the ultimate foundation upon which the subjective temporal flow is based. Varela's argument here seems based on little more than a verbal analogy centering on the expression 'self-driven' (Varela, 1999, pp. 128–130). The sequence of transient phase-locking episodes is self-driven in that it occurs as a function of the intrinsic dynamics of the neural oscillators and their coupling, and does not require any outside influence. The absolute time-constituting flow (recall the discussion in Sect. 2) is 'self-driven' in the sense that it is taken to be a structure of phases of temporal contents that explains time-consciousness without itself being 'located in', or flowing in, any distinct conscious temporal flow. Poetic license aside, if the proposal were correct, then differences in the time courses of these patterns should be undetectable by the subject. But they are detectable. So they can't be the neural substratum of the absolute time-constituting flow.

### 3.4 Cerebral blood flow

The most recent, and most sophisticated, proposal to be discussed comes from Dan Lloyd (Lloyd, 2002). Lloyd focuses not on dynamical systems theory or coupled oscillators, but on temporal sequences of fMRI data. The goal is to discern in patterns in how brain states, as revealed by fMRI, change over time. And the hope is that some of what is found will be "... analogues of phenomenal structures, particularly the structures of temporality." (Lloyd, 2002, p. 818)

The first study targets the 'temporal flux' of time consciousness, the idea that there is a psychologically real flow to time such that, among other things, even if one is repeating a certain task, one will nevertheless be aware of the fact that time has passed between these tasks. The guiding thought of the first study is that "[t]ime is the river that carries all else, so a strong prediction would be that the flux of time would appear as a monotonic increase of intervolumetric multivariate difference as a function of intervening interval in time, or lag between images." (Lloyd, 2002, p. 821)

Recall from Sect. 2 the recursive nature of now-consciousness: each new 'now' includes an awareness of the previous 'now' as something that has just past, and the next 'now' will likewise include an awareness of the present 'now', including its own inclusion of the prior 'now', in itself. There is thus something like a

recursive accumulation of awareness-of-nows—“... a nested cascade of prior states ...” (Lloyd, 2002, p. 825). *If* our awareness of the passage of time involves this sort of structure, and *if* this structure is conceived on analogy with a sort of accumulation (it need not be, other metaphors can be imagined, including ‘nested cascade’), and if one then moves without fanfare from content to vehicle, one can produce a neuro-physiological prediction: there is some sort of accumulation in the neural processes, or more mathematically, some sort of strictly monotonic change in brain states over time. If we are lucky, these will be detectable by fMRI. Lloyd develops techniques for analyzing large corpora of fMRI data, and these techniques reveal that there are detectable strictly monotonic changes over time in the brain. More specifically, if one compares the fMRI-assessed state of the brain at two different times, one will find smaller differences between states separated by smaller temporal intervals than one will find if one compares states across larger temporal intervals.

This result, however, is a nearly trivial though fairly non-obvious consequence of the way that the data is processed. Each image in the series consists of some number of voxels—volume elements analogous to pixels in a two-dimensional image. Each voxel in each image has a scalar value. The Euclidean distance between two images is arrived at as the sum of the squares of all voxel value differences. For example, suppose that we have two images (‘1’ and ‘2’) each with three voxels (‘A’, ‘B’ and ‘C’) on each image. And let’s define notation such that  $A_1$  is the scalar value of voxel A on image 1. Then the ‘distance’ between the image 1 and image 2 is

$$d(1,2) = (A_1 - A_2)^2 + (B_1 - B_2)^2 + (C_1 - C_2)^2$$

Roughly, Lloyd’s analysis shows that as  $n$  increases,  $d(1,n)$  increases monotonically—that is, as the sequence of voxel images progresses, the distance between that image and the first image increases. What might this monotonic change be caused by? Imagine that you put a leaf on the surface of a lake. You might imagine that small random buffets might move the leaf around, but if they are random, they would *not always* move the leaf away from the spot where you dropped it—it would move sometimes farther, sometimes backtrack. If the leaf was found to consistently move farther from the initial spot at each step, you would take this to indicate some sort of current, perhaps one you had not noticed before. Lloyd found just such a consistent increase in distance, and takes himself to have found the current of the river that carries all else.

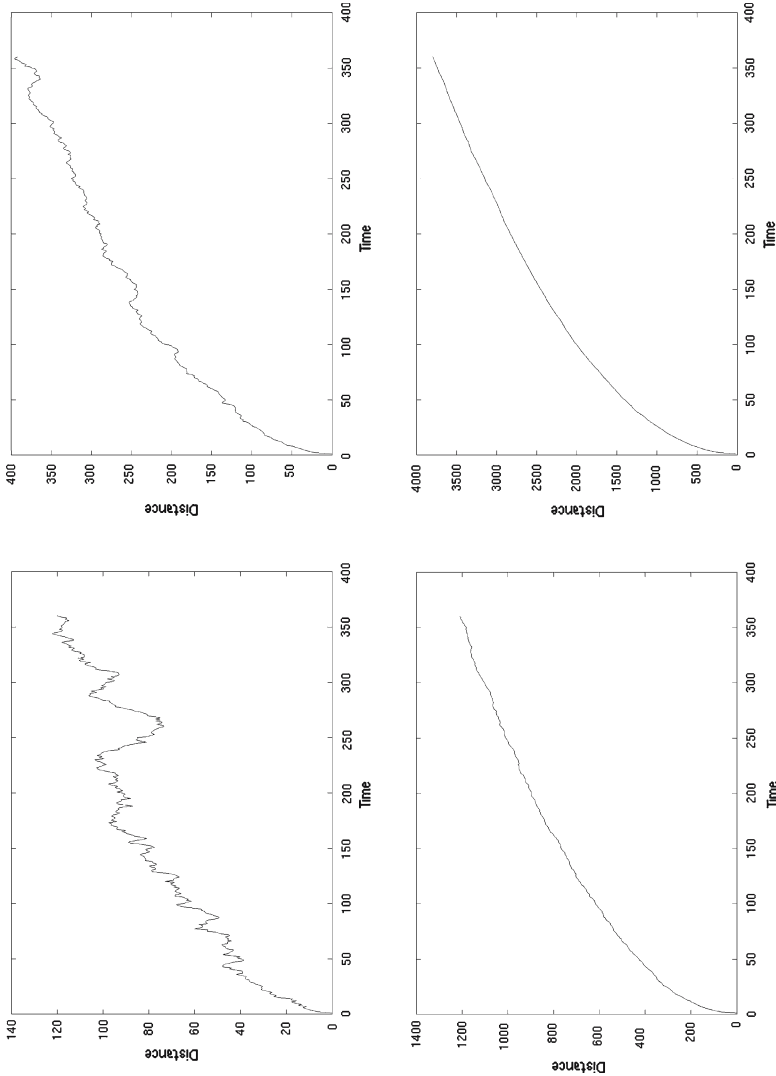
Unfortunately, he has found no such thing. The intuitions about the leaf on the lake have their home in spaces of one to three dimensions. But as the dimensionality of a space increases, many of these intuitions become increasingly misleading. In particular, the intuition to the effect that random noise would not always result in an increase in distance is false in spaces of large dimensionality. Suppose we have only one voxel (and hence one dimension), whose value is pushed around by small Gaussian noise. The distance  $d(1,2)$  (between the voxel’s value on image 1 and image 2) will be a result of this noise. The subsequent image, image 3, may move a bit further away, or may move closer to where it was on image one. Since we have only one dimension, there is a 50/50 chance on each time step of the distance increasing or decreasing. And so there is no reason to expect a monotonic increase in distance. But what happens when we move to two voxels? The values of the second image will differ from the first, as each voxel’s value is nudged by the noise. Suppose half increase from 0 to 5, and half decrease from 0 to  $-5$ . Here  $d(1,2)$  is 50. What happens on the 3rd image? Again let’s suppose each value is nudged either farther or closer randomly. If they

are both nudged farther, to 10 and  $-10$ , then  $d(1, 3)$  will increase to 200; and if both are nudged closer, back to 0 and 0, then  $d(1, 3)$  will decrease to 0. But, and here is the crucial point, if *one* moves closer and the *other* farther—to either (0, 10) or  $(-10, 0)$  the net distance  $d(1, 3)$  will still *increase*, to 100. That is, when we move from one to two dimensions, the chances of a decrease in Euclidean distance on subsequent images drops from 50% to 25%, and the chances of an increase in distance go from 50% to 75%. Why? Because of the way Euclidean distance is measured. Because Euclidean distance is based on a sum of *squared* distances, any *increase* in the value of one of the dimensions has a larger positive effect on the resultant distance than an equal decrease in the value of another dimension has a negative effect. Because of this, as the number of dimensions increases, the odds that random noise on any step will result in a decrease in Euclidean distance get smaller and smaller. To put it equivalently, as the number of voxels increases, the odds that random noise will fail to result in a strictly monotonic increase in Euclidean distance get diminishingly small.

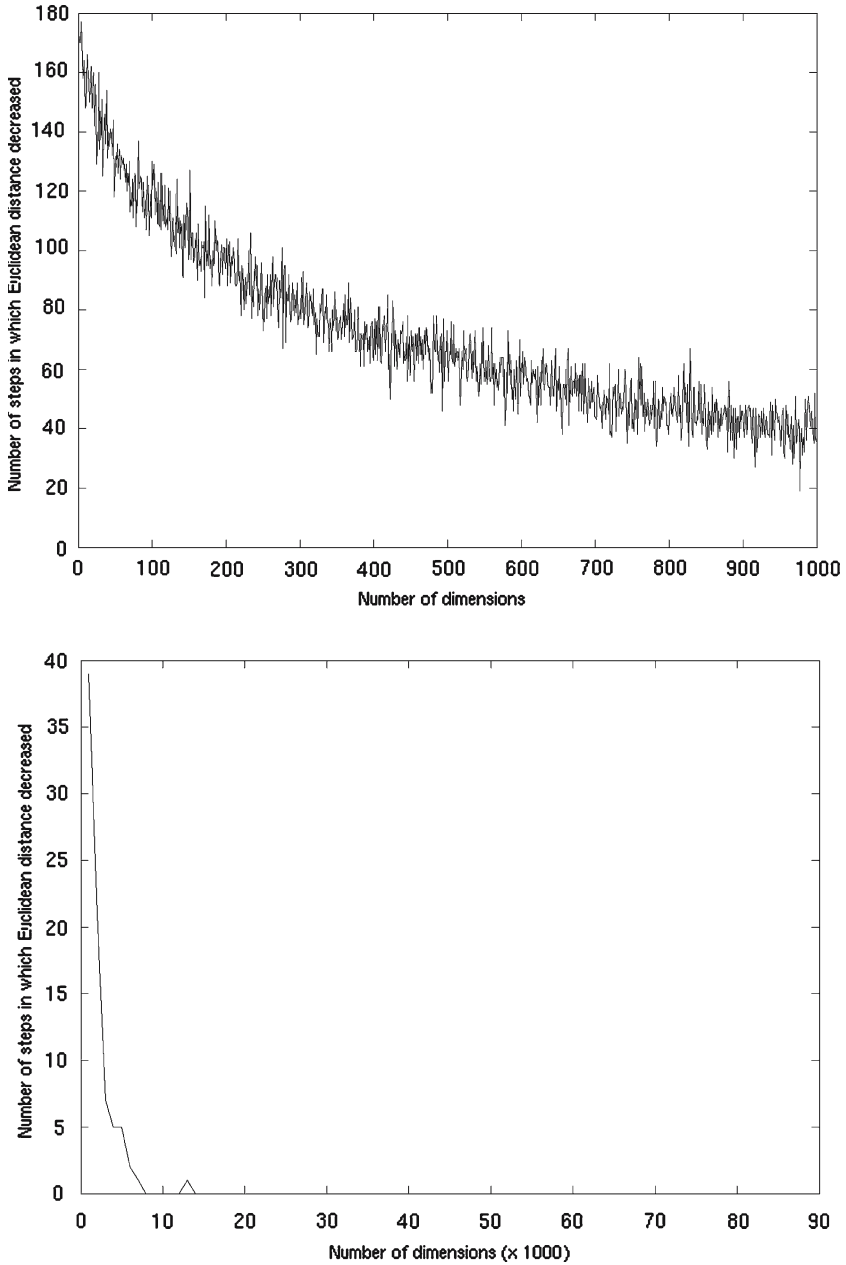
This is illustrated in Figs. 1 and 2. The graphs in Fig. 1 were constructed by taking sequences of artificial voxel images and plotting the Euclidean distance of the images in the sequence from the initial image. Each image in an artificial sequence consists merely of the previous image plus a very small zero-mean additive Gaussian noise added to each voxel value. That is, *there is absolutely nothing happening in the image sequence except random noise*, and yet the sequences exhibit exactly the monotonic increase in distance that Lloyd's analysis exhibited in the real fMRI data. As can be seen from the graphs in Fig. 1, as the number of voxels in a given image sequence goes up, the effect becomes more uniformly monotonic. This is illustrated in Fig. 2, which plots the number of noise-generated successor images, out of a series of 360, that produce an image *closer* to the initial image than its predecessor, as a function of the number of voxels. For image sequences using only a few voxels (the left of the first graph), about 150 out of 360 steps result in a distance decrease. As the number of voxels goes to 1000, the number of distance-reducing steps drops to about 40. The second graph of Fig. 2 plots the same thing, from 1000 voxels to 86,000 voxels, in increments of 1000 (86,000 was chosen because it is approximately the number of voxels in the study by Postle et al. (2000), a study that Lloyd used for his analysis). As can be seen, once the number of voxels reaches 14,000, there was *never* a step that resulted in a decrease in Euclidean distance.

So the detection of a monotonic increase in Euclidean distance between voxel images in a series would appear to be explainable, at least in principle, by nothing more interesting than random noise, the Euclidean metric, and the large number of dimensions involved in the images. Dan Lloyd (personal communication) has pointed out that

But what you've done to simulate your time series is demonstrably not happening in the brain. Voxel values in the brain always hover about their mean value. The added noise in real images is not accumulative. To see that your simulation is unrealistic, just plot the standard deviations of the images in your series. It will increase steadily from image to image. Standard deviation in a real image series is essentially constant. (Or just plot your voxel time series to observe their drift away from their starting values.) Real image series are not accumulating noise; real voxels are not on a random walk. So your conjecture does not explain the observation.



**Fig. 1** Four plots of Euclidean distance as a function of 'time' for artificial voxel sequences subject only to additive Gaussian noise. The upper left plot is for a sequence of 360 images where each image consists of 10 voxels. The upper right is for images of 1000 voxels. And the lower right is for images of 10,000 voxels. In all cases, distance increases with time. As the number of voxels increases, the distance increase becomes more smoothly monotonic



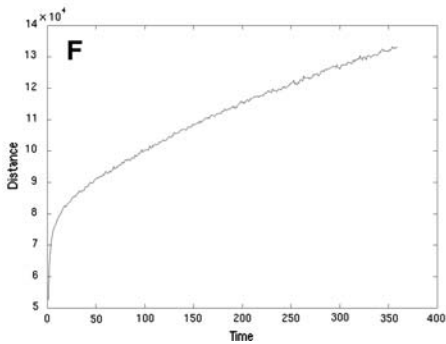
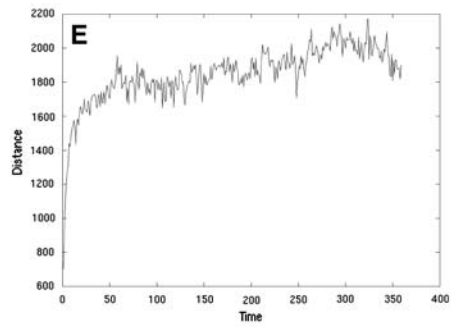
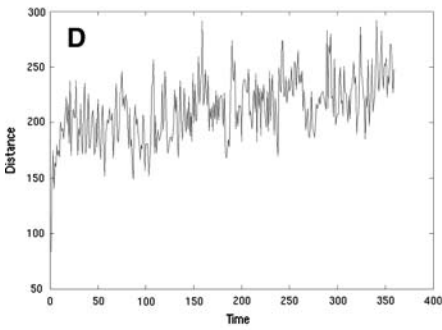
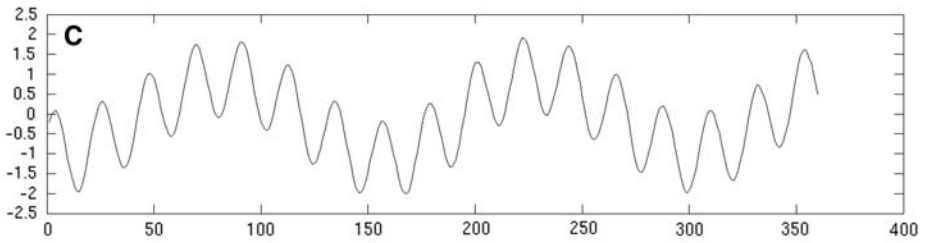
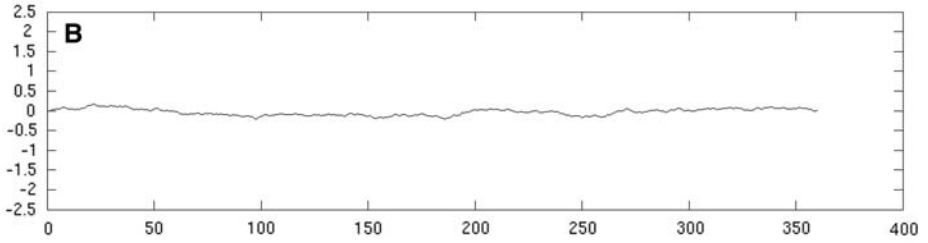
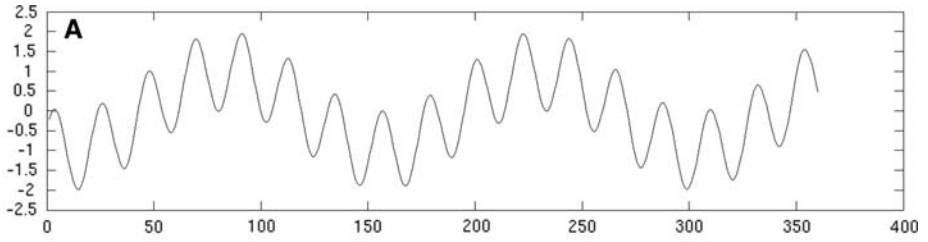
**Fig. 2** The number of decreases in Euclidean distance, out of 360 time steps, that result from random noise, as a function of the number of dimensions (voxels) in the space. The plot on the top shows number of decreases for spaces of 1 to 1000 dimensions, where the number of decreases drops from about 160 to around 40. The plot on the bottom shows decreases for higher dimensional spaces, from 1000 to 86,000. The number of decreases drops from 40 at 1000 to 0 at 14,000

There are points here that are right, and some that I believe are misguided. It is true that these simulations allowed the voxel values to go on a random walk. That was the point, not to provide what must be an explanation of the phenomenon (who knows what the explanation really is?), but rather to point to one trivial thing that could explain it. But if in fact the voxel values do ‘hover about their mean’ then it must be a special sort of hovering about a single value that is consistent with an overall monotonic distance increase over time. The distance increase over time is Lloyd’s own finding, not mine. Given this finding, the values can’t *simply* be hovering around any single value. Simply hovering about any value would not result in a monotonic distance increase, by definition. At best there is something that looks very much like hovering but with an overall distance increase. Hovering about a mean supplemented by nothing more than very small additive Gaussian noise would do the trick. This is illustrated in Fig. 3.

But at risk of beating a dead horse, it deserves to be pointed out that regardless of what the explanation for this result is, the result fails to lead to any interesting conclusions about time consciousness, since it is neither necessary nor sufficient for anything related to Husserlian analyses of time consciousness. That they are not sufficient can be seen by noting that any theory of the mind that holds that the *mind* changes monotonically over time will, if it is allowed to exploit the same metaphors and content/vehicle slides, make exactly the same prediction. According to Locke, for example, the mind is a sort of storehouse of ideas that it experiences. As we experience things, the associated ideas are stored by the mind, and these ideas are then available to be dredged up again. And even if over some period no new ideas are presented, new *associations* between ideas will be formed. Locke, and in particular aspects of Locke’s theory that are not at all concerned with time-consciousness, posits a monotonic increase in a number of mental items and states over time, and so a neuro-Lockean, if allowed access to the same metaphors and content/vehicle slides, would make exactly the same “strong prediction” about a monotonic change in detectable brain states. And Locke’s theory of mind is, by Husserl’s lights, one that is manifestly incapable of explaining time consciousness! So a monotonic change in brain states is not sufficient for anything distinctively Husserlian. So a monotonic change is consistent with Husserl’s program in exactly the way that it is also consistent with programs (e.g., Locke’s) that Husserl’s program is a denial of.

That changes in fMRI-detectable brain states are not *necessary* for time consciousness is also an easy point. First off, there are plenty of brain states that are not fMRI detectable, and so even if time-consciousness is in fact a matter of some sort of monotonic change in some brain state, this state need not be one that is detectable by fMRI. Second and more importantly, even if the relevant brain states are ones that are fMRI

**Fig. 3** Illustration of a second set of simulations. Each artificial voxel’s value was a sum of two factors. ► First, a sum of two sine waves, each with amplitude 1, and random phase and frequency. This results in a value that hovers around a mean value, in this case swinging from 2 to  $-2$ . The second factor is a relatively tiny amount of additive Gaussian noise (B). In these simulations, the noise sigma was .02, only *one one-hundredth* the magnitude of the voxel’s range of values as determined by the two sine waves. The sum is shown in (C), which is nearly indistinguishable from (A). (D), (E) and (F) are plots of distance from the initial voxel values as a function of time for 100 voxels, 1000 voxels, and 86,000 voxels, respectively. As can be seen, the same result holds: as the number of voxels increases, noise — even an amount of noise relatively minuscule compared to the range of each voxel’s variance — results in an overall monotonic distance increase



detectable, the relevant change need not be one that is monotonic. The discussion of binary representations from Sect. 3.1 is sufficient to make the point. If time is tracked by ticking up a binary clock (and who is to say it isn't?), the relevant distance measures between the physical states at different times, as assessed by capacitor charges or their neural analogues, will not change monotonically, but will fluctuate up and down. More generally, it is quite possible for time-representing mechanisms to be non-monotonic, in that increases in time are represented, but not by monotonic-increasing vehicle properties. To summarize my criticisms of Lloyd's first study: first, the monotonic change in brain state that was detected is possibly nothing but an artifact of the distance measure used, the number of dimensions involved, and random noise; second, *even if there were* some non-trivial monotonic change in brain state, this wouldn't have any direct bearing on the relation between the brain and Husserlian analyses of time consciousness.

A second study by Lloyd is meant to provide a sort of neuro-vindication of Husserl's tripartite analysis of time consciousness. The content-level phenomenon focused on is the fact that on Husserl's analysis, protention, primal impression, and retention are not contentfully unrelated, but rather are contentfully related in specific ways—namely, protention is an anticipation of imminent primal impression, and retention is retained primal impression. Thus the complete content grasped at any moment is related to the content grasped at a nearby moment, it is "... a superposition of the object's history and possible future" (Lloyd, 2002, p. 825). This was tested by seeing if a neural network would be able to accurately reproduce successor and predecessor fMRI data when given as input the data from a given time. As Lloyd describes it

Phenomenologically, each moment of consciousness is a sandwich of past, present, and future. Accordingly, each pattern of activity in the brain will be inflected with past and future as well. But "past" and "future" can only be understood internally, that is, as past and future states of the brain. To discover tripartite temporality, then, we seek to detect some form of continuous neural encoding of past states, as well as some anticipation of the future. (Lloyd, 2002, p. 821)

This was tested in the following way. Normal voxel series were processed into a small number of principle components: orthogonal dimensions of variation such that the first dimension is the dimension that captures the greatest variation in the series, the next component is the orthogonal dimension that captures the most of the remaining variance, and so forth. A neural net was trained whose job was to reconstruct either the successor or predecessor principle component representation of a voxel images when given a principle component representation of a 'present' voxel image as input. This network achieved a certain level of success. A second network was trained on a surrogate set of data that was produced in the following way. A discrete Fourier transform was effected on each principle component time series. The phases were then shuffled, and the inverse Fourier transform implemented to produce a set of principle component time series that was in some ways statistically similar to the original series (auto-correlation within a principle component time series was retained), but such that correlations between the different principle component time series was destroyed.

The network's performance on these surrogate voxel series was significantly less than its performance on the real voxel series. But the fact that a network performed better on the real series than on the surrogate series is not, as far as I can tell, surprising at all. What would have to be the case in order for the real and surrogate

data to allow for equally good performance? Well, since the only difference is that the real series, but not the surrogate series, retains correlational information between *different* principle components in temporally adjacent images, it would have to be the case that only significant correlation is autocorrelation. Under what conditions would autocorrelation be the only correlation that mattered? When a brain state corresponding to a principle component depended causally on only the that specific principle component's own prior state, not on the prior state of any other components.

So what we can conclude from the study is that this view is wrong. The brain is in fact *not* composed of a large number of causally isolated microvolumes or principle components. That is, different regions of the brain causally interact. I'm pretty sure that this has been the accepted view in neuroscience for some time. I'm also pretty sure that this says nothing particularly interesting about time consciousness. As far as I can tell, the suggestion that it does is contained in the following line of Lloyd's reasoning:

Controls for each probe suggest that the probe network performance depends on more than simple serial correlation in voxel time series, and on something other than general statistical profiles of the training and test images. This suggests that the brain encompasses a distributed encoding of its own past and future. That past and future brain state information is embedded in present brain states is consistent with the phenomenological claim that retention and protention are superposed in the conscious awareness of the subjective present. (Lloyd, 2002, pp. 827–828)

The similarity between this and van Gelder's remark that the past is 'built in' to the state of a dynamical system in virtue of its causal priority is striking. The 'distributed encoding of its own past and future' is in fact no more than the causal fan in and fan out of various brain regions. The solar system has exactly such a distributed encoding of its own past and future. I agree that it is true that this is *consistent* with the phenomenological claim. But for my own part, I agree because I have a hard time imagining any sane view of brain function with which it is not consistent.

### 3.5 Gray codes, convexity, and isomorphisms

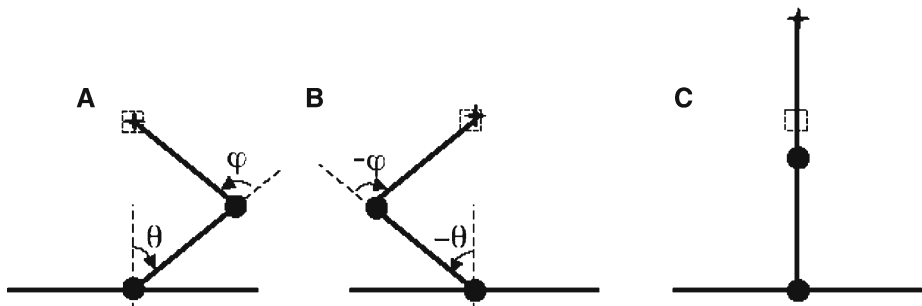
Recall again the points raised in Sect. 3.1 about binary representations of numbers and Hamming distances. Situations where a small or unit change in the number represented by a code results in a large Hamming distance, or vice versa, is a *Hamming cliff*. The metaphor is clear enough—one small step in Hamming distance results in a huge difference in the number represented, or vice versa. Standard binary representations, as well as normal decimal representations, are full of Hamming cliffs. They present often-unrecognized challenges to, e.g., connectionist networks where inputs are represented as binary input vectors, and genetic algorithms where small changes in the binary code of an artificial genome should not generally yield completely unrelated phenotypes. Encoding schemes specifically designed to avoid Hamming cliffs, such that a unit difference in the number represented always corresponds to a unit Hamming distance are known as *Gray codes*, and are of great practical use in many areas, including connectionist models and genetic algorithms.

Given this, someone, especially Lloyd, might object to what was said in Sect. 3.1 along the following lines: "Sure, some artificial codes might have these odd properties that foil the move from vehicle properties to content properties, but natural ones

don't. In particular, neural representation formats are very likely convex (see, e.g., Gärdenfors, 2004), meaning that the average of any elements in the set is itself an element of the set. From convexity something analogous to a Gray code for neural representations can be derived. And with this in hand, the move from similarities of representational vehicle to similarities of represented content follow." Or similarly, one might simply argue directly that neural representation is generally a matter of isomorphism (see e.g., O'Brien & Opie, 2006).

There are several points to make in this regard. First, many domains dealt with by natural systems, and the associated neural representations, are not convex. Kinematics—the relation between joint angles and resultant body posture—is one example that provides for a very simple illustration of the point. Consider a toy example of a single arm with a shoulder and elbow joint, and the goal of grasping a target, as illustrated in Fig. 4. Here, the goal is obtained by shoulder angle  $\varphi$  and elbow angle  $\theta$ . As illustrated in B, the set of angles  $-\varphi$  and  $-\theta$  is also a solution to this problem. Both of these are elements of the solution set of the kinematic problem. But note that the average of these two kinematic solutions is not itself a kinematic solution. The average of  $(\varphi, \theta)$  and  $(-\varphi, -\theta)$  is obviously  $(0, 0)$ , and this is not a solution to the kinematic problem. Domains in which a sensorimotor system has redundant degrees of freedom—and this is the rule in motor control, which is as biologically plausible a domain as one could hope for—are typically also such that they are not convex. So the convexity/Gray code gambit, at least in this form, falters.

The point of these considerations is to drive home the fact that one cannot just assume without argument, or without even addressing the issue, either that (i) properties of contents carried by vehicles can be read off any of the physical properties of the vehicles; or that (ii) relations between contents carried by sets of vehicles can be read of relations between physical properties of the vehicles. Of course, it is open to one to explicitly take this task up and demonstrate, or at least provide some plausibility to the suggestion, that in the cases at hand the required relation between content and vehicle obtains, as O'Brien and Opie (2006) and Gärdenfors (2004) have (though they do not discuss temporal representation specifically). I have my doubts about the



**Fig. 4** A toy illustration of the fact that kinematics is not a convex domain, in that the average of two solutions to a kinematic problem need not itself be a solution to that same problem. The figures illustrate simple arm with a shoulder and elbow, and the kinematic problem is to determine joint angles that will get the 'hand' (indicated by a 4-point star) to the target (indicated by a small dashed-line box). The solution illustrated in (A) is shoulder angle  $\theta$  and elbow angle  $\varphi$ . The solution illustrated in (B) is shoulder angle  $-\theta$  and elbow angle  $-\varphi$ . As illustrated in (C), the average of these two solutions, shoulder angle 0 and elbow angle 0, is not itself a solution to the problem

prospects for any such program, especially for the case of temporal representation. But the point is that none of the accounts I have criticized in this paper have, so far as I can tell, even recognized the content/vehicle issue as an issue, let alone provided any reason to follow the needed isomorphism assumption. You don't solve a problem by failing to recognize it.

#### 4 On processing temporal information

The fact that content/vehicle confusions are confusions does not entail that *all* properties of the vehicles of a representation are irrelevant for explaining the content carried by that representation. A legitimate part of the explanation of why a given inscription means *blue* might well appeal to features of the inscription itself—not its color, of course, but the relative arrangement of the curves and line segments that compose the letters of the inscription. Granted this is not the entire explanation—appeal would need to be made to the norms of the language community, and perhaps much else. But some properties of the vehicle can be explanatorily relevant to the content they carry. Indeed, it would be hard to imagine a case where *all* vehicle properties were irrelevant. But the point is that this relevance cannot simply be *assumed* to be a matter of isomorphism or iconicity or any other similarly simple coding scheme. In short, while there must be some route from vehicle properties to content properties, this route is not necessarily, and in most cases is not, direct (isomorphism, iconicity), but rather *indirect* in one way or another.

I believe that features of the neural information processing machinery in the central nervous system *are* relevant to those representational structures that underwrite the temporal aspects of our conscious experience.<sup>7</sup> Furthermore, I believe that, to a first approximation at least, Husserl's analysis does accurately characterize certain aspects of our subjective experience. And since I also believe that our phenomenal experience is largely a function of the representational structures produced by neural information processing machinery, I am committed to there being something about the mechanisms neural *information processing* that explains why our phenomenal experience explains those features of phenomenology revealed by Husserl's analysis.

What is needed to do the job responsibly is a middle-level theory that explicitly addresses the issue of how content properties, so to speak, are implemented in vehicle properties. We can't just jump between features of vehicles to conclusions about features of contents. Conveniently enough, I have elsewhere developed a theoretical framework for understanding the information processing structure of the temporal aspects of the perceptual system that is up to the task: the trajectory estimation model (my discussion here will be very brief, please see Grush, 2005a, 2005b for more detail).

The trajectory estimation model is based upon, and is a sort of generalization of, internal modeling approaches that focus on state estimation. The basic idea of internal modeling approaches is that the system has an internal model of the perceived entity (typically the environment and entities in it, but perhaps also the body), and at

<sup>7</sup> They are relevant, but not *necessarily* for reasons analogous to the inscription case. In the case of inscriptions, the vehicle properties are properties that an interpreter discerns in order to begin the interpretation process. In the case of neural mechanisms, analogous reasoning would yield something like a homunculus or other interpreter. While I think that something like this is close to correct, the point is that from the hypothesis that vehicle properties are relevant it does not follow that they are relevant because of their role in enabling 3rd party interpretation.

each time  $t$ , the state of the internal model embodies an estimate of the state of the perceived domain. The model can be run off-line in order to produce expectations of what the modeled domain might do in this or that circumstance (or if this or that action were taken by the agent); the model can also be run online, in parallel with the modeled domain, in order to help filter noise from sensory signals, and in order to overcome potential problems with feedback delays (see Grush, 2004a, 2004b for a review of many such applications and references).

We can formalize the basic idea with some simple notation. Let  $p(t)$  be a vector of values that specifies the state of the represented domain (or at least those aspects of it that are relevant for the representing system). And for simplicity let's assume that the system is a driven Gauss–Markov process, meaning that its state at any time is determined by four factors: its previous state; the laws (e.g., laws of physics) that determine how the state evolves over time; a driving force, which is any predictable influence on the system; and process disturbance, which is any unpredictable disturbance. In equation form

$$p(t) = Vp(t - 1) + d(t) + m(t) \quad (1)$$

where  $p(t)$  is the process's state vector;  $V$  is a function that captures the regularities that describe how the process evolves over time;  $d(t)$  is a driving force, which is any *predictable* influence on the process's state; and  $m(t)$  is a small zero-mean additive Gaussian vector that represents any *unpredictable* influence on the process's state, sometimes called *process noise* (though I prefer *process disturbance* since it is a real effect, the expression 'noise' erroneously suggests to many that it is not a real effect).

At each time  $i$ , the controlling or cognitive system produces an estimate  $\hat{p}(i)$  of the state of the process as it is at time  $i$ . One common strategy for producing this estimate is to combine knowledge of how the process typically behaves with information about the process's state provided by sensors. Formally, this can be described as follows. First, the system uses its previous state estimate together with its knowledge of the predictable driving force, and its knowledge of the regularities that describe how the process evolves over time—that is, knowledge of  $V$ —to produce an a priori state estimate

$$\bar{p}(t) = V\hat{p}(t - 1) + d(t) \quad (2)$$

Here,  $\hat{p}(t - 1)$  is the previous state estimate,  $V$  is the function describing how the state typically evolves over time, and  $d(t)$  is the driving force. This a priori estimate,  $\bar{p}(t)$ , will be accurate only to the extent that the previous estimate was accurate, and will also not take into account the process disturbance  $m(t)$ , since it is unpredictable. The second factor used to construct the estimate is information about the process's state provided by noisy sensors. At all times a noisy signal  $s(t)$  is produced that can be conceived as a noise-free measurement of the process—produced by a measurement function  $O$ —to which non-additive Gaussian noise  $n(t)$  is added

$$s(t) = Op(t) + n(t) \quad (3)$$

This factor does not depend on the accuracy of any previous estimates, nor is it foiled by process disturbance, since such disturbance really does affect  $p(t)$ , and since  $p(t)$  is what is measured, the observed signal  $s(t)$  captures information about the effects of process disturbance. However, this factor is subject to sensor noise. Combining the

two factors allows for a better estimate than is possible from either individually

$$\hat{p}(t) = \bar{p}(t) + kO^{-1}(O\bar{p}(t) - s(t)) \tag{4}$$

Here  $\hat{p}(t)$  is the final, a posteriori state estimate. It is arrived at by taking the a priori estimate  $\bar{p}(t)$  and adding a correction term, which is derived from the difference between what the observed signal actually is, and what the observed signal was expected to be ( $O\bar{p}(t)$ ). The gain term  $k$  determines the relative weight given to the sensory information and the a priori estimate in forming the a posteriori estimate.

This anyway is the basic idea behind many internal modeling approaches. For much more detail, including many applications to visual imagery, motor imagery, visual processing, motor control, and even a few remarks on neural mechanisms, see Grush (2004a, 2004b).<sup>8</sup>

Now to the trajectory estimation framework, which is a generalization of this approach, according to which the system maintains, at all times  $i$ , not an estimate of the process's state at  $i$ , but an estimate of the trajectory of the process over the temporal interval  $i - l$  to  $i + k$ , for some relatively small temporal durations  $l$  and  $k$ . To streamline the notation, let  $\hat{p}_{h/i}$  be the estimate, produced at time  $i$ , of the state of the process as it is/was/will be at time  $h$ . This notation can be generalized to  $\hat{p}_{[i-j, i+k]/i}$ , which is an estimate, produced at time  $i$ , of the behavior of domain  $p$  throughout the temporal interval  $[i - j, i + k]$ . It will be convenient for some purposes to describe this in discrete terms, as an ordered  $j + k + 1$ -tuple of state estimates  $(\hat{p}_{i-j/i}, \dots, \hat{p}_{i/i}, \dots, \hat{p}_{i+k/i})$ .

In terms of information processing, largely the same mechanisms that are able to produce current process state estimates can be employed to produce the estimates of the other, past and future, phases of the trajectory estimate. Predictions of predictions of a priori predictions of future states of the process in the obvious way

$$\bar{p}(t + 1) = V\hat{p}(t) + d(t + 1) \tag{5}$$

And this process can obviously be iterated to produce, at time  $i$ , estimates of what the process's state will be at any arbitrary future time  $i + k$ , so long as knowledge of  $d(i + k)$  is available.

Estimates of previous states of the process can be arrived at via smoothing

$$\tilde{p}(t - 1) = \hat{p}(t - 1) + h(V^{-1}\hat{p}(t) - d(t)) \tag{6}$$

Here, the smoothed estimate  $\tilde{p}(t - 1)$  is arrived at by adding to the filtered estimate  $\hat{p}(t - 1)$  a correction term based on the filtered estimate from the subsequent time step. Here,  $V^{-1}$  is the inverse of the function  $V$  that maps current to successive process states, and so  $V^{-1}\hat{p}(t)$  is the expected predecessor state to  $\hat{p}(t)$ , where here 'expected' means 'modulo driving force and process disturbance'; and  $h$  is a gain term. Equation 6 can obviously be applied recursively to produce estimates of the state of the process at time  $i - j$  for arbitrary lag  $j$

$$\tilde{p}(t - 2) = \hat{p}(t - 2) + h(V^{-1}\tilde{p}(t - 1) - d(t - 1)) \tag{7}$$

One way to maintain a trajectory estimate then is to just maintain at all times an estimate of the related set of state estimates, estimates for states of the process from  $i - l$  to  $i + k$ . A qualitative description of state estimation and trajectory estimation

<sup>8</sup> For those who feel that internal modeling approaches ignore the insights of equilibrium-point and similar models of motor control (e.g. Balasubramaniam, 2004; and Latash & Feldman, 2004, both commentaries to Grush, 2004a), see my reply in Grush (2004b), Sect. R3.

would be as follows. A *state* estimator uses knowledge about how the process typically behaves over time, together with information in the observed signals up to time  $i$ , to produce at time  $i$ , an estimate of the state of the process at time  $i$ . A trajectory estimator uses the same sources of knowledge, knowledge of how the system typically behaves over time and the observed signals up to time  $i$ , but uses them to a different purpose—it uses them to produce, at time  $i$ , an estimate of the process's behavior over a temporal interval. This trajectory estimate includes as one aspect an estimate of the current state of the process. But it also includes improved estimates of the recent prior states<sup>9</sup>, as well as anticipations of imminent states.

That anyway is a schematic description of the information processing framework. There is reason to think that this information processing structure is actually implemented by the human perceptual system, with a lag and reach on the order of 100 ms each, for a total temporal magnitude on the order of 200 ms. It will be convenient to address the past-oriented phase first. Getting at this phase is tricky, since what is sought is evidence to the effect that the human perceptual system is maintaining at any moment a representation not just of the state of the perceived domain at that instant, but rather of a temporal interval about 100 ms of the domain's behavior. It might be tempting to dismiss this possibility on the grounds that it does not seem as though we are seeing 100 ms worth of motion at an instant. A bowling ball looks like a moving bowling ball, and not like an irregular cylinder whose length is the distance the ball travels in 100 ms. This line of thought presupposes that the temporal interval is represented as something like a time-lapse photograph, which of course is not how it would be represented according to the trajectory estimation model. A time-lapse photograph represents all the phases within the 100 ms interval as simultaneous. But on the trajectory estimation model, the prior phases are represented as being prior, as things that *just happened*. And so according to the trajectory estimation model, your perceptual experience of the bowling ball should present it as being at a current location now, but as having just been at a slightly different location just prior to that, and so forth. And it is not clear that this is in conflict with the phenomenal facts.

Indeed, one line of evidence, and one historically appealed to in support of the idea that the temporal contents of perception comprehend a temporal interval is exactly the fact that we can perceive motion. Motion can only be manifested over a temporal duration, and so if we can perceive motion (as opposed to always merely inferring motion), then perceptual experience must comprehend a temporal interval.

The line I find most compelling focuses on temporal illusions—cases where subjects are mistaken about the temporal features of things they perceive. For example, Geldard and Sherrick (1972) found that a certain sort of illusion could be induced by tactile stimuli. The experimental setup involved placing small mechanical devices at various places on subjects' arms and shoulders. These would produce sequences of small taps, the exact nature and timing of these sequences under the control of the experimenters. Some of the sequences lead to no surprising results: a sequence of taps all located at the same spot on the wrist, for example, will be reported by the subject as a sequence of taps at the same location at the wrist. However, different sequences provide more interesting results.

<sup>9</sup> Smoothed estimates are typically improvements over the corresponding filtered estimates. The filtered estimate that was produced at time  $t-l$  of the process's state at time  $t-l$  took into consideration sensor information up to time  $t-l$ . The smoothed estimate of the process's state at time  $t-l$ , produced at time  $t$ , takes into account sensor information collected up to time  $t$ .

“... if five brief pulses (2-msec duration each, separated by 40 to 80 msec) are delivered to one locus just proximal to the wrist, and then, without break in the regularity of the train, five more are given at a locus 10 cm centrad, and then another five are added at a point 10 cm proximal to the second and near the elbow, the successive taps will not be felt at the three loci only. They will seem to be distributed, with more or less uniform spacing, from the region of the first contactor to that of the third.” (Geldard & Sherrick, 1972, p. 178)

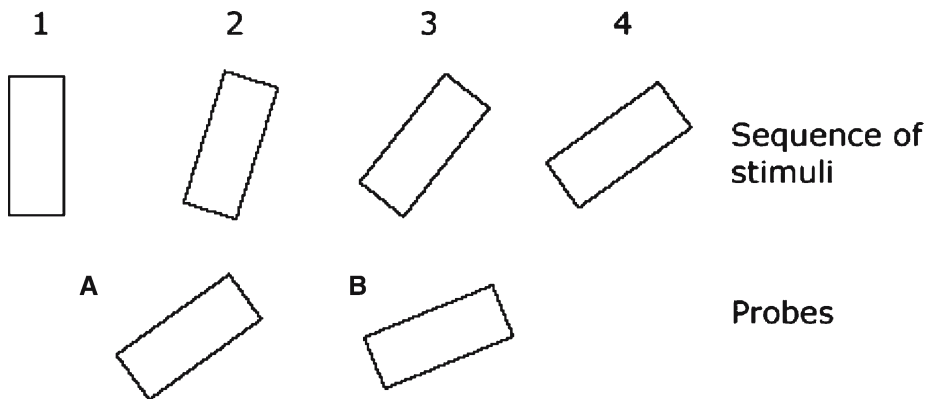
This is not merely a spatial illusion, but a temporal one as well, as can be brought into focus by asking where the subject feels the second tap at the time of the second tap.

If the human perceptual system is implementing something like a trajectory estimation model, then the following would be a description of what is happening. *At the time of the second tap*, the content of the perceptual experience places the second tap at the wrist. The observed signal indicates that this is what happened, and there is nothing about that signal that is odd. However, at some point, which Geldard and Sherrick were getting at experimentally, the perceptual system decides that it is more likely that the observed signals of taps at three discreet locations is more likely to be an inaccurately sensed series of evenly spaced taps than an accurately sensed set of taps at three locations. This of course requires that there is some knowledge about the statistics of the environment indicating what sorts of patterns are likely and what sorts are unlikely.

Similar illusions manifest in other modalities. An example is apparent motion, where a sequence of two flashes in close spatial and temporal proximity is seen as a single moving dot. Until the second dot flashes, there is no way to know whether there will be a second flash, nor, if there is, which direction it will be in. Yet the moving dot is seen as being at the intermediate locations before being at the terminal location. There is evidence that a temporal interval on the order of 100 ms or so is a maximum, in the visual case anyway, for retrodiction effect. This implies that the trajectory whose estimation supplies the content of perceptual experience spans an interval into the past on the order of a hundred milliseconds or so (for more detail, see Grush, 2004b).

The considerations I find most compelling for the existence and magnitude of the future-oriented aspect of the trajectory estimator is representational momentum. The original Geldard and Sherrick article briefly mentions, like an afterthought and without further exploration, that “there is typically the impression that the taps extend beyond the terminal contactor” (Geldard & Sherrick, 1972, p. 178). This effect—the apparent continuation of some perceived stimulus motion beyond its actual termination—has been studied a great deal under the rubric of *representational momentum*. A typical stimulus set together with its perceived counterpart are shown in Fig. 5.

While there are many possible explanations for this phenomenon, it certainly suggests that at some level the perceptual system produces representations whose content anticipates, presumably on the basis of the current observations and past regularities, the immanent antics of the perceived situation. And the temporal magnitude appears to be on the order of 100 ms or so. In this context it is interesting to note that the representational momentum effect appears to be tied to predictability (Kerzel, 2002). It is true that the phenomenon is most often introduced with examples involving the apparent continuation of linear or circular motion, but cases that are significantly more complicated also exhibit the phenomenon so long as they are predictable.



**Fig. 5** Representational momentum. A sequence of stimuli is shown to subjects, such as a moving ball or a rotating rectangle. The sequence is ended by a masking stimulus. Subjects are then shown two probe stimuli, such as two different end locations for the rectilinear motion, or rectangles oriented at different angles for the rotating motion, and are to select the one that matches the last stage of the movement that they observed. Subjects overshoot by preferring probes that slightly overshoot the actual terminus to those that accurately mirror the terminus. For review see Thornton & Hubbard (2002)

Perhaps the most interesting is the highly nonlinear case of biomechanical motion (Verfaillie & Daems, 2002).

This concludes the brief introduction to the trajectory estimation model. The two crucial features of the trajectory estimation model for present purposes are (i) that it is dealing with representational contents, and (ii) it is capable of relatively straightforward neural implementation. As to the first point, the trajectory estimation model is *not* a description of any physical states of a system, nor of any states of idealized units that are taken to correspond to physical states of a neural implementation, like firing rates or oxygen consumption. It is an information processing description, meaning that it specifies how a system that represents certain kind of information in certain kinds of formats can manipulate this information in order to arrive at certain kinds of structures of representations. It is thus at least a contender for comparison with Husserlian phenomenology—*itself* a theory of the structure of perceptual content, and not a theory of neural firing rates or oxygen consumption.

As to the second point, I have said nothing about physical implementation here. The model is silent on implementation. This can seem odd for a proposal that I have advertised as a possible bridge between Husserlian phenomenology of time consciousness and computational neuroscience. Let me simply point out that the model does allow for unmysterious implementation. Anyone interested in how are invited to take a look at Eliasmith & Anderson (2003); Haykin (2001); and a recent special issue of the *Journal of Neural Engineering*, devoted to internal modeling approaches (Poon & Merfeld, 2005). None of these sources discuss trajectory estimation per se, but they do discuss neural net implementation of control and filtering models from which the trajectory estimation model is constructed.

## 5 Discussion and conclusion

I will close this paper by first by discussing a few respects in which the trajectory estimation model fails to fully correspond to features of Husserl's program, and

second by pointing out a major commonality between my own view and those I have criticized in this paper.

While there are some obvious similarities between the trajectory estimation model and Husserl's analysis of time consciousness, there are some significant mismatches. First, Husserlian protention and retention do not seem to be limited to intervals on the order of 100 or 200 ms. Husserl's examples of melodies clearly indicate that he had ranges *at least* on the order of seconds in mind. And to the extent that there is a genuine phenomenon present in, e.g., music perception, as there appears to be, a few hundred milliseconds seems far short of the magnitude required.<sup>10</sup>

However, there are actually several phenomena to be discerned here that are not adequately separated by Husserl or the researchers who have followed him (or who have followed the related Jamesian doctrine of the specious present). There are two ways in which it could be plausibly maintained that contents characterizable only in temporal interval terms play a role in experience. One, which potentially spans a larger interval, might be described as *conceptual* in the sense that it is a matter of interpreting present experience in terms of concepts of processes that span potentially large intervals. Music appreciation would fall into this category. When I recognize something as part of a larger whole (a spatial whole or a temporal whole), then my concept of that whole influences the content grasped via the part. Something along these lines is what appears to be happening with music. On the other hand, there is what might be called a perceptual or phenomenal phenomenon of much brief magnitude. In the music case, the listener is quite able to draw a distinction between some things she is perceiving and some she is not, and notes from a bar that sounded three seconds ago will not typically be misapprehended by the subject as being currently perceived, even though their presence is felt in another, contextual or conceptual sense.

By contrast, the representational momentum and perceptual retrodiction phenomena are cases where the it *does* seem to the subject that she is perceiving, the relevant content. The point is easiest to make in the case of perceptual retrodiction. When the subject perceives the dot as having moved from point A to point B, she has no recollection of having perceived anything different. How this phenomenon gets characterized (perceptual versus memory) is not relevant. What is relevant is that whatever this phenomenon is, it is a matter of the contents grasped by the subject at a time that concern temporal processes, and that it is limited to brief intervals on the order of a few hundred milliseconds. So any mismatch on this score between Husserlian analyses and the trajectory estimation model do not indicate that one or the other is wrong, but rather that there are at least two phenomena here, two different kinds of retentional-protentional structure in play, and Husserl focuses on the one active over longer durations, and the trajectory estimation model is an attempt to explain the one active over shorter durations. Though I should say that Husserl is not nearly as clear on this topic as one might hope.

Second, it might be objected that Husserl's account is supposed to be a phenomenological account, and what I have offered is at best an information processing account. My reply to this is that while Husserl uses the expression 'phenomenology', his analysis is not about qualia, or what to a modern ear might be suggested by 'phenomenal content' or anything like that. Rather, Husserl's analysis is pitched almost entirely at the intentional level, that is, as an analysis of the contents grasped in experience.

<sup>10</sup> Dan Lloyd has suggested that in part because of this the trajectory estimation model is perhaps a better model of a Jamesian specious present doctrine than Husserlian analyses.

And this is precisely what the trajectory estimation model is a model of: structures of representational contents.

Third, Husserl's analysis involves more than just retention, protention and primal impression. There are additional *exotica* and *obscura* such as the absolute time-constituting flow, and its 'double intentionality', and the recursive structure of the content of each of the grasped now-phases. The trajectory estimation model is not addressing any of these phenomena. Worse, at least for the rhetorical aims of this paper, some of the competitors discussed in Sect. 3 do concern themselves with some of these phenomena. Lloyd makes appeal to the recursive structure of the content of various temporal phases, and Varela takes the absolute flow to be addressed by the dynamical properties of cell assemblies. And so it might seem as though the trajectory estimation model has a shortcoming that some of its competitors lack. My response to this should be no surprise. If the competition addressing these phenomena had anything remotely revealing to say about them, then this would be a relevant point in their favor. But they don't. So it's not. Theories don't gain adequacy points through addressing 'additional' phenomena in manifestly inadequate ways. The trajectory estimation model does not address these additional phenomena, and so in that sense it is incomplete as a bridge to a full Husserlian program.

The fourth and fifth points of disanalogy are such that the right conclusion seems to me to be that Husserl's analysis is flawed. These two points are related and derive ultimately, I believe, from a residual Cartesian hangover on Husserl's part. First, Husserl privileges a now-point and gives it the name *primal impression*. The trajectory estimation model does not privilege any of the phases within the temporal interval. Part of the difference here is explained by the fact that, as mentioned above, each seems to be addressing a related but different phenomenon operative at different time scales. At the larger scale, singling out a portion as present and as phenomenally quite unlike the earlier and later phases seems right. At the smaller time scale, however, this may not be a legitimate move. Of course within the 200 ms interval temporal discriminations are made, earlier and later phases are temporally distinguished—that is, grasped as earlier and later. But this does not require that any of the phases is singled out as present, with all others as future or past.

The fifth point is that Husserl sees the relation between protention, retention and primal impression as one of 'modifying' items that remain in other respects constant. This is suggested by the name 'retention' itself, but is also explicitly stated as a feature of the analysis. On the trajectory estimation model this is not what happens. As time progresses, the *entire trajectory* is re-estimated, with the consequence that some parts of the estimate can be changed. For example, according to the trajectory estimation model, in the cutaneous rabbit situation, at the time of the second tap the relevant part of the trajectory estimate is 'second tap at the wrist'. If Husserl were correct about the way that retention operates, then this estimate should simply sink back, unchanged but for its temporal marker, as time progresses. But as we have seen, this need not happen. At some point, if the sequence of stimuli is right, the trajectory estimate will be modified so that the relevant retention will have the content 'there was a second tap *proximal to the wrist*'. And this will be the correct explication of the content of that retention, at that time, even though there never was a primal impression with that content. On Husserl's analysis, temporal illusions should not be possible. But they are possible, so Husserl's analysis can't be right in this respect. So much for the points of disanalogy between the trajectory estimation model and Husserl's analysis.

It should be obvious enough that while I have been highly critical of van Gelder, Varela and Lloyd, there is a clear sense in which the four of us are on the same team. We all believe that an important source of insights for the task of understanding of mentality is what Lloyd describes as ‘analytic phenomenology’, even if we disagree about how to go about harvesting these insights. But some may wonder why worry about this. Why all the hubbub about Husserl? The tradition Husserl was a part of, and that I, van Gelder, Varela and Lloyd take seriously, has as a central task the discovery of general principles of mentality, or conscious experience. I could say quite a bit about this, but a few remarks will have to suffice. The congenitally blind are not mindless, so clearly vision is not a necessary feature of a mind. And those who are able to recall more, or fewer, items from a list than is normal are also not lacking a mind. Nor is it clear that emotions necessary to have or be a mind. A person who for whatever reason lacks a capacity to experience emotions might be interesting, in that she might not be able to do some things as well as people who can experience emotions. Perhaps, even, her reasoning will be impaired. But impaired reasoning is not the lack of a mind. It is, rather, a mind that is less able than usual to do some sort of task.

The questions *What is a mind? What would some entity have to have, or be able to do, in order for it to be or have a mind?* are, interestingly, questions that are simply not raised by the sciences of the mind. That these sciences are oblivious to their own lack of concern about discerning general principles about their presumed object of study is surprising. I have invariably been met with puzzled looks when I raise such questions to psychologists or neuroscientists, and it takes a little time and effort to get them to see what the question is! Reassuringly, they all eventually understand the question, and often agree that it is an interesting and possibly important one, even if it is one they simply had never even thought about. But like I said, this is not the place to explore this issue. The point for now is that the tradition Husserl was part of was one that took such questions seriously. The commonality between myself and those I have criticized is that we take that this task is an important one, one that can aid, and be aided by, empirical investigations as carried out by the relevant sciences. This is no trivial commonality. It is, in my opinion, ultimately of far greater theoretical importance than the differences that I have focused on in this paper.

**Acknowledgements** A version of this paper was presented to the UCSD Philosophy Fight Club, and benefited from feedback at that session. I also received excellent comments and suggestions on a prior version of this paper from Dan Lloyd, Gualtierro Piccinini, and an anonymous referee for this journal. Lloyd deserves special mention for providing thoughtful and unusually detailed and useful comments despite the fact that the paper is highly critical of one of his projects. I would also like to thank the McDonnell Project in Philosophy and the Neurosciences, and the Project’s director Kathleen Akins, and acting director Martin Hahn, for grant support.

## References

- Andersen, S. (1994). A computational model of auditory pattern recognition. Doctoral Dissertation, University of Indiana.
- Balasubramaniam, R. (2004). Redundancy in the nervous system: Where internal models collapse. *Behavioral and Brain Sciences*, 27(3), 396–397.
- Brough, J. (1989). Husserl’s phenomenology of time-consciousness. In J. N. Mohanty, & W. R. McKenna (Eds.), *Husserl’s phenomenology: A textbook* (pp. 249–290). Lanham, MD: University Press of America.

- Clark, A. (2000). *A theory of sentience*. Oxford: Oxford University Press.
- Eliasmith, C., & Anderson, C. (2003). *Neural engineering: Computational, representation, and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.
- Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. Cambridge MA: MIT Press.
- Geldard, F. A., & Sherrick, C. E. (1972). The cutaneous “Rabbit”: A perceptual illusion. *Science*, 178(4057), 178–179.
- Grush, R. (2004a). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(3), 377–396.
- Grush, R. (2004b). Further explorations of the empirical and theoretical aspects of the emulation theory. *Behavioral and Brain Sciences*, 27(3), 425–442.
- Grush, R. (2005a). Brain time and phenomenological time. In Brook, & Akins (Eds.), *Cognition and the brain: The philosophy and neuroscience movement* (pp. 160–207). Cambridge: Cambridge University Press.
- Grush, R. (2005b). 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), S209–S218.
- Haykin, S. (2001). *Kalman filtering and neural networks*. New York: Wiley.
- Husserl, E. (1991). (translation by John Brough of Husserl (1966)). *On the phenomenology of the consciousness of internal time (1893–1917)*, *Collected works IV*. Dordrecht, Boston, and London: Kluwer Academic Publishers.
- Husserl, E. (1966). *Zur Phaenomenologie des inneren Zeitbewusstseins [1893–1917]*, Herausgegeben von Rudolf Boehm, *Husserliana X*, The Hague.
- James, W. (1890). *Principles of Psychology*. New York: Henry Holt.
- Kerzel, D. (2002). A matter of design: No representational momentum without predictability. *Visual Cognition*, 9, 66–80.
- Latash, M. L., & Feldman, A. G. (2004). Computational ideas developed within the control theory have limited relevance to control processes in living systems. *Behavioral and Brain Sciences*, 27(3), p. 408.
- Lloyd, D. (2002). Functional MRI and the study of human consciousness. *Journal of Cognitive Neuroscience*, 14(6), 818–831.
- O’Brien, G., & Opie, J. (forthcoming). How do connectionist networks compute? *Cognitive Processing*, 7.
- Postle, B. R., Berger, J. S., Taich, A. M., & D’Esposito, M. (2000). Activity in human frontal cortex associated with spatial working memory and saccadic behavior. *Journal of Cognitive Neuroscience*, 12, 2–14.
- Poon, C.-S., & Merfeld, D. M. (Eds.). (2005). Sensory integration, state estimation, and motor control in the brain: Role of internal models. *Special Issue of the Journal of Neural Engineering*, 2(3).
- Sorensen, R. (2004). We see in the dark. *Nous*, 38(3) 456–480.
- Thornton, I. M., & Hubbard, T. L. (2002). Representational momentum: New findings, new directions. *Visual Cognition*, 9(1/2), 1–7.
- Van Gelder, T. (1996). Wooden Iron? Husserlian Phenomenology Meets Cognitive Science, *Electronic Journal of Analytic Philosophy*, 4.
- Varela, F. (1999). Present-time consciousness. *Journal of Consciousness Studies*, 6(2–3), 111–140.
- Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in Vision. *Visual Cognition*, 9, 217–232.