J. Neural Eng. **2** (2005) S209–S218 doi:10.1088/1741-2560/2/3/S05

Internal models and the construction of time: generalizing from *state* estimation to *trajectory* estimation to address temporal features of perception, including temporal illusions

Rick Grush

University of California, San Diego, CA, USA

Received 11 February 2005
Accepted for publication 1 July 2005
Published 31 August 2005
Online at stacks.iop.org/JNE/2/S209

Abstract

The question of whether time is its own best representation is explored. Though there is theoretical debate between proponents of internal models and embedded cognition proponents (e.g. Brooks R 1991 *Artificial Intelligence* **47** 139–59) concerning whether the world is its own best model, proponents of internal models are often content to let time be its own best representation. This happens via the time update of the model that simply allows the model's state to evolve along with the state of the modeled domain. I argue that this is neither necessary nor advisable. I show that this is not *necessary* by describing how internal modeling approaches can be generalized to schemes that explicitly represent time by maintaining trajectory estimates rather than state estimates. Though there are a variety of ways this could be done, I illustrate the proposal with a scheme that combines filtering, smoothing and prediction to maintain an estimate of the modeled domain's trajectory over time. I show that letting time be its own representation is not *advisable* by showing how trajectory estimation schemes can provide accounts of temporal illusions, such as apparent motion, that pose serious difficulties for any scheme that lets time be its own representation.

1. Introduction

The deepest theoretical divide in the sciences of cognition and its physical bases is between, on the one hand, theories that highlight the role of internal representations—paradigmatically internal models—of the agent's body and environment in explaining an agent's behavior, and, on the other hand, theories that highlight the role of high-bandwidth agent–environment interactions in producing adaptive behavior without much or any representation on the part of the agent. The debate is not new. The behaviorism of the early 20th century was in large part a reaction against explanation in terms of internal mental states when it was felt that explanations in terms of organism responses to environmental stimuli would be explanatorily sufficient and more theoretically parsimonious (Watson 1913, Skinner 1935). The early cyberneticists (Wiener 1948, Ashby 1952) brought concepts from classical control theory and dynamical systems theory to bear on behavior explanation by highlighting representationless feedback mechanisms and dynamical agent–environment interactions. The 'cognitive revolution' (Tolman 1948, Chomsky 1959) argued that such mechanisms, while perhaps more parsimonious, were in fact not explanatorily sufficient for a wide variety of cognitive phenomena, most notably learning. More recently, Rodney Brooks has given fresh inspiration to the anti-representationalists by building robots that manage to complete various tasks without any internal models, but by *letting the world be its own model*, as Brooks put it (Brooks 1991).

One of Brooks' main arguments hinges on the claim that representations are bottlenecks in perception–action cycles.

Brooks conceives representations as being located in a series that includes sensation modules and action modules. But, of course, internal models need not have this character. A Kalman filter, for example, runs an internal model *in parallel with*, not in series with, the perception–action cycle. The Kalman filter is thus an example of representation without bottlenecks. See Grush (2003) for more discussion of the impact of schemes that employ internal models on the representationalism debate.

But there is one respect in which even the staunchest proponent of internal models has been in unwitting complicity with the anti-representationalist—*time*. While internal models of the body or environment are not content to let the body or the environment be its own representation, such models nevertheless allow time to be its own representation. It is common to assume that the domain being modeled evolves over time in regular ways, paradigmatically and most tractably as a driven Gauss–Markov process, where the state of the process at any time is determined by three factors: a successor function of its previous state; a predictable driving force; and unpredictable disturbances, often modeled as additive Gaussian noise. Accordingly, the model of the domain is assumed to exploit knowledge of the successor function that governs the evolution of the modeled domain. This 'time update' is automatically applied to the model in such a way as to allow it to evolve over time just as the modeled domain's state automatically evolves over time. This is one of the primary factors that allow the model's state to be a reliable estimate of the modeled domain's state as time progresses.

But note that in such cases time is being used as its own representation. Such a system represents a temporal feature such as succession, say that state A of the environment occurred before state B of the environment occurred, via the temporal features of the representations—the *representation of A* occurred before the *representation of B*. The representation of succession is accomplished by a succession of representations, and can only be so represented on such schemes. To put it another way, the temporal features of the mechanisms that do the representing is exploited to represent the time of the domain that is represented. Time is represented with time. It is silently assumed to be its own best model, indeed its only model.

In section 2, I will briefly describe one way in which a scheme that utilizes internal models could do so in a way that does not just let time be its own representation, but rather actively and explicitly represents the temporal aspects of the modeled domain just as typical modeling schemes model the nontemporal aspects. Briefly, the proposal is for the internal modeling not of temporally punctate states of the modeled domain, but of the trajectory of the modeled domain over a temporal interval.

While section 2 tries to show that there are theoretical options available besides letting time be its own representation, section 3 addresses the issue of why it might be beneficial to do so. After all, why not join in with the anti-representationalists on this one issue? It certainly seems like time is special in that both the model and the modeled domain are subject to the exact same physical laws, and are both evolving through time in such a way as to track each other accurately, assuming that they are not moving at some substantial fraction of the speed of light with respect to each other, anyway. One reason, though not the only one, is the existence of temporal illusions. Generally speaking, illusions are a strong motivation for positing representations, since a representational theory has an easy explanation of illusions as the production of representations that do not accurately represent the modeled domain. The anti-representationalist has a hard time with illusions. If the world is being used as its own representation, how can it look to be other than it in fact is? Temporal illusions present the same challenge to the anti-representationalist. If time is its own representation, then it would seem that temporal illusions ought not to be possible.

In section 3, I will discuss a number of temporal illusions, including the flash-lag effect (MacKay 1958), the cutaneous rabbit (Geldard and Sherrick 1972) and apparent motion (Kolers 1972). As something of an appetizer, I will briefly describe the *cutaneous rabbit* now. The basic phenomenon is this: a test subject has small tactile stimulators placed on her arm, one near the wrist, a second between the wrist and elbow, and a third near the elbow. If the mechanical stimulators produce a sequence of five taps at the wrist, each separated by 40–80 ms, then the subject will report feeling a group of taps near the wrist. But if the sequence of taps produced is five taps near the wrist, followed by five at a location between the wrist and elbow, and finally five at the elbow, the subject will report feeling an evenly spaced sequence of taps progressing from the wrist to the elbow. Initially this can seem like a merely spatial illusion, since what is being inaccurately represented in the second case is the location of many of the taps. The second tap of the second sequence is felt not on the wrist, but a few centimeters proximal to the wrist. But there is a puzzling temporal aspect to the phenomenon as well that can be brought out by reflecting on the question: Where does the subject feel the second tap when the second tap is produced by the stimulator? The answer seems to be the following: if in the immediate future there will be more taps only at the wrist, then the second tap will be felt at the wrist; but if in the future an appropriate sequence of taps will be delivered to the forearm and elbow, then the second tap will be felt a few centimeters proximal to the wrist. But at the time of the second tap, the subject cannot know what the future sequence of taps will be, since the different possible sequences are randomly selected. Surely the perceptual system cannot look into the future, see where the subsequent taps will be delivered and use that information to decide how to interpret the location of the current tap!

In section 4, I will discuss how trajectory estimation schemes of the sort I described in section 2 can address these phenomena. Part of the discussion will include a comparison of the trajectory estimation approach to two other approaches: Dennett and Kinsbourne's *multiple drafts model* (Dennett and Kinsbourne 1992), and Rao, Eagleman and Sejnowski's *smoothing model* (Rao *et al* 2001). In a final discussion section, I will briefly point to some additional areas of application of the trajectory estimation model.

2. Trajectory estimation

The goal of this paper is a conceptual goal. I am not going to articulate a particular concrete model and then compare the performance of that model to empirical data, or anything of the sort. Rather, my goal is to bring to light a wider range of salient options to those who wish to apply internal models to perceptual processing, and to point out some of the advantages of exploring some of those options. All of the conceptual points I wish to make can be illustrated while restricting the discussion to discrete linear systems. For simplicity, I will assume that the modeled domain is a driven Gauss–Markov process:

$$p(t) = Vp(t-1) + d(t) + m(t). \tag{1}$$

Here, $p(t)$ is an $n \times 1$ vector describing the process's state; V is a function, represented as an $n \times n$ matrix, that maps states of the process onto successor states of the process; $d(t)$ 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*.

The process's state is measured at each time by one or more sensors. This measurement is presumed to be noisy. The production of the noisy observed signal can be formally described as a noise-free measurement of the process's state to which a small zero-mean time-dependent non-additive Gaussian noise vector is added. The noise-free measurement can be represented as a function O that maps process states onto signal states:

$$I(t) = Op(t). \tag{2}$$

Here $I(t)$ is the real, noise-free sensory signal at time t. The *observed signal* $s(t)$ is a noisy version of $I(t)$:

$$s(t) = I(t) + n(t). \tag{3}$$

Again for simplicity, I will assume that V and O are both invertible.

As an exemplar of an internal modeling approach to controlling the process, or simply filtering noise from the observed signal, consider a scheme that maintains an estimate of the state of the process by exploiting knowledge of the function V that maps process states to successor states, the function O that effects a measurement of process states to signals, and has access to the observed signal at each time step. The model will exploit a prediction–correction cycle as follows. First, an *a priori* estimate of the process' state is produced by iterating the estimate from the previous cycle and adding the predictable driving force:

$$\bar{p}(t) = V\hat{p}(t-1) + d(t). \tag{4}$$

Here $\bar{p}(t)$ is the *a priori* state estimate, and $\hat{p}(t-1)$ is the *a posteriori* estimate from the previous estimation cycle. This *a priori* estimate is measured to produce an *a priori* estimate of the signal:

$$\bar{I}(t) = O\bar{p}(t) \tag{5}$$

This *a priori* signal estimate is compared to the observed signal and the difference, the *sensory residual*, is pushed through a measurement inverse. The result is multiplied by a gain to yield a correction to the *a priori* estimate:

$$\hat{p}(t) = \bar{p}(t) + kO^{-1}(\bar{I}(t) - s(t)). \tag{6}$$

Here $\hat{p}(t)$ is the final *a posteriori* state estimate, and k is a gain that determines the relative weight given to the sensory residual and the *a priori* estimate in forming the *a posteriori* estimate.

Finally, if the system is filtering the signal, then the final *a posteriori* signal estimate $\hat{I}(t)$ is given by

$$\hat{I}(t) = O\hat{p}(t). \tag{7}$$

For review of many applications of this sort of approach to understanding various functions of the nervous system, see Grush (2004).

The same mechanisms and information can be used to produce *a priori* predictions of future states of the process in the obvious way:

$$\bar{p}(t+1) = V\hat{p}(t) + d(t+1). \tag{8}$$

And this process can obviously be iterated to produce, at time t, estimates of what the process's state will be at any arbitrary future time $t+k$, so long as knowledge of $d(t+k)$ is available; the availability of future intentions will be significant for a specific purpose discussed in section 4.

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

$$\tilde{p}(t-1) = \hat{p}(t-1) + h(V^{-1}\hat{p}(t) - d(t)). \tag{9}$$

Here, the smoothed estimate $\tilde{p}(t-1)$, which is produced at time t, is arrived at by adding to the filtered estimate $\hat{p}(t-1)$, which was produced at time $t-1$, a correction term based on the filtered estimate from the subsequent time step t. 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 (9) can obviously be applied recursively to produce estimates of the state of the process at time $t-j$ for arbitrary lag j:

$$\tilde{p}(t-2) = \hat{p}(t-2) + h(V^{-1}\tilde{p}(t-1) - d(t-1)). \tag{10}$$

With such tools in place, it is possible to describe a system that combines smoothing, filtering and prediction to maintain an estimate of the trajectory of the modeled domain over the temporal interval $[t-j, t+k]$, by determining, at each time t, the following ordered $j+k+1-$ tuple:

$$(\tilde{p}(t-j), \tilde{p}(t-j+1), \ldots, \hat{p}(t), \bar{p}(t+1), \ldots, \bar{p}(t+k)). \tag{11}$$

Since following how trajectory estimates themselves change as time progresses will be a central theme in what follows, it will be convenient to have, on each state estimate that is a component of a given trajectory estimate, *separate indices* for the time at which that estimate is produced and the time that the estimate is representing. So I will streamline the notation by letting $\hat{p}_{i/h}$ be the estimate produced at time h of the state of the process at time i, and it will be understood that if $i < h$ the estimate is smoothed, if $i = h$ it is filtered and if $i > h$

it is predicted. Similarly, I will extrapolate to notation of the form $\hat{p}_{[f,g]/h}$ for a trajectory estimate produced at time h of the process's trajectory from $t = f$ to $t = g$ inclusive.

Before moving on to a discussion of temporal illusions, there are two aspects of a trajectory estimation scheme of the sort I have sketched here that I wish to emphasize. First, time is not being used to represent time. It is possible to represent a succession without a succession of representations, for example. A single trajectory estimate, produced at one time, is capable of representing temporal relations of various sorts, e.g. succession, simultaneity, duration, without needing to use time to represent these relations.

The second point is obvious, but I will highlight it anyway. The system is continually updating its estimate of the entire trajectory as new measurements come in at each time. As a result, the estimate of a state of the process at a given time might change. There is no reason to expect that $\hat{p}_{r/s,}$ as an element of $\hat{p}_{[s-j,s+k]/s}$, will be the same as $\hat{p}_{r/s+1}$, as an element of $\hat{p}_{[(s+1)-j,(s+1)+k]/s+1}$ (for $s+1-j \leqslant r \leqslant s+k$), even though both are estimates of the process's state at time $t = r$.

3. Application to temporal aspects of perceptual processing

3.1. The puzzles of the cutaneous rabbit and apparent motion

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:

> ... 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 central, 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 and Sherrick 1972, p 178)

I explained in the introduction why this was not merely a spatial illusion, but presented a temporal puzzle as well, one brought into focus by asking where the subject feels the second tap at the time of the second tap.

The phenomenon of apparent motion presents exactly the same paradox. Two successive flashing dots presented within some spatial distance and within some inter-stimulus interval will appear to be a single moving dot, moving from the location of the one that flashes first to the location of the one that flashes second (see figure 1). Again, this can look to be merely a spatial illusion in that it looks as though a dot has moved through spatial areas where no dot has in fact been—it appears as though the dot occupied and moved through location B as indicated on the right-hand side of figure 1. To bring out the temporality of the phenomenon, consider that the subject will appear to see the dot first at location A, then at location B, and finally at location C—the motion is actually perceived to be continuous, but I am just drawing attention to the temporal relations between three of the positions on the continuous path.

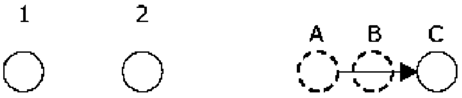

Figure 1. Apparent motion. The left-hand side represents actual stimuli, a flashing dot (1) followed by a second flashing dot (2). The right-hand side represents what is perceived: a single dot moving from location A (the location of the first dot's flash), through location B and to location C (the location of the second dot's flash).

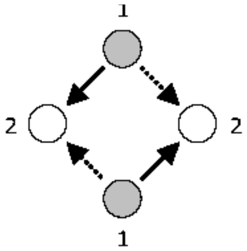

Figure 2. An ambiguous bi-stable quartet of flashing dots. First, the upper and lower dots (labeled '1') flash simultaneously, and then the left and right dots (labeled '2') flash. The resulting apparent motion is either clockwise (dotted-line arrows) or counterclockwise (solid-line arrows).

Note, however, that if the second flashing dot were above, below or to the left of the first, then the subject would have seen the dot as moving upward, or leftward or downward. And accordingly, the intermediate location B would be either above, to the left of or below location A. But, and this is the crucial bit, until the second dot actually flashes, the subject cannot know in which of these four spatial directions the interpolated motion (the location of B) should occur. Yet the subject sees the dot as being at the interpolated location before being at the terminal location where the second flash occurs. It can seem as though the perceptual system is able to foretell where the second flash will be in order to appropriately begin filling in the intermediary phases of the apparent motion.

An even more interesting apparent motion phenomenon has recently been studied (Williams *et al* 2005). Consider a bi-stable quartet of flashing dots, as in figure 2 (Gengrelli 1948). The resulting apparent motion will be seen either as (a) the top dot moving down and to the left, with the bottom dot moving up and to the right (counterclockwise motion indicated by the solid-line arrows); or (b) the top dot moving down and to the right, with the bottom dot moving up and to the left (clockwise motion indicated by the dotted-line arrows). A diamond orientation bi-stable quartet is perceived as clockwise motion about half the time, and as counterclockwise about half the time (Ramachandran and Anstis 1983).

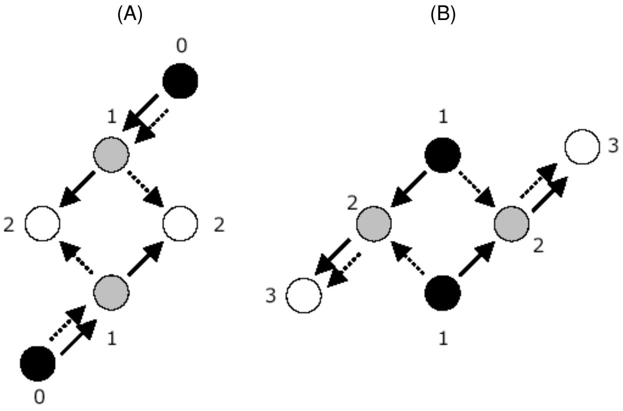

Figure 3. Apparent motion retrodiction. See the text for explanation.

The left-hand side of figure 3 shows how the perceived direction of motion of the bi-stable quartet can be influenced by prior flashing dots. In the situation diagramed, before the four dots in the quartet flash, a third pair outside the quartet is flashed. The sequence of dot flashes could yield two possible apparent motion sequences. First, the motion from the two external dots to the top and bottom dots on the quartet could be followed by counterclockwise motion of the quartet, yielding two perceived rectilinear apparent motions as indicated by the solid-line arrows. Second, the motion from the external dots to the top and bottom dots could be followed by clockwise motion of the quartet, yielding two perceived left-hand-turn movements as indicated by the dashed-line arrows. Perhaps not terribly surprisingly, the visual system has an apparent preference for rectilinear motion (Ramachandran and Anstis 1983). This suggests that the visual system employs an internal model of object motion according to which rectilinear motion is significantly more likely than non-rectilinear, and the invocation of this motion model by the observed signals corresponding to the first and second pair of dots biases the system to perceive one path of motion rather than the other. This is a fascinating result, but not entirely unexpected by anyone who takes it that the perceptual system is set up to exploit knowledge of the sorts of processes that are likely in order to produce anticipations of what will be observed next.

The phenomenon illustrated on the right-hand side of figure 3 is yet more interesting. The first two pairs of dots that flash are identical to the pairs that flash in the ambiguous bi-stable quartet. However, after these dots, a third pair flashes. As in the previous case, there are two possible paths of apparent motion: a rectilinear path (indicated by the solid-line arrows), and a left-hand-turn path, indicated by the dashed-line arrows. Since it is known that the bi-stable quartet will produce the clockwise and counterclockwise apparent motions with equal probability, it would be expected that the rectilinear path and the left-hand-turn path should be perceived with equal probability. However, the result is that when the successive pairs of dots followed each other at 67 ms, rectilinear motion was perceived significantly more than 50% of the time (see Williams *et al* (2005) for details, including the strength of the effect, and how the strength varies as conditions are modified; for trials in which the interstimulus interval was 100 ms, no significant effect was found).

3.2. *The solution to the 'paradoxes'*

The trajectory estimation model applies to the cutaneous rabbit as follows. At $t = 1$ and $t = 2$, the observed signals are two taps on the wrist. At $t = 2$ the trajectory estimate will simply be that two taps have been felt at the wrist. So to the question: At the time of the second tap, where does the subject feel that tap? The answer is: *at the time of the second tap*, the second tap is felt at the wrist. This is true regardless of what the future sequence of taps will be. If we suppose that the trajectory estimate produced spans ten time steps before and after the present time, then we can indicate this by saying that $\hat{p}_{2/2}$ (an element of $\hat{p}_{[-8,12]/2}$) is 'tap on the wrist'.

However, suppose that in the immediate future, say from $t = 3$ to $t = 10$, taps are delivered to locations proximal to the wrist as in the Geldard and Sherrick experimental condition. The trajectory estimator will, if the sequence is right, be forced to make a choice between interpreting the sequence of taps as being either (i) three groups of taps delivered to spatially discrete locations that were *accurately reported* by the sense receptors; or (ii) an evenly spaced series of taps from the wrist to the elbow that was *inaccurately reported* by the sensors. For some sequences of stimuli the second trajectory estimate is produced. Presumably this is because the nervous system has models of external objects and their likely trajectories, embodied in an analogue of V that is appropriate to environmental stimuli, which indicates that continuous motion is more likely than discontinuous motion. In such a case, $\hat{p}_{2/10}$ (an element of $\hat{p}_{[0,20]/10}$) is 'tap just proximal to the wrist'. If the system is probed at $t = 10$ or later, it will 'report' that it has just observed a sequence of evenly spaces stimuli. So even though *at the time of the second tap* it is felt at the wrist, if the system is given an appropriate sequence of subsequent taps and is probed late enough, it will *then* represent the second tap *as having been* just proximal to the wrist.

A similar explanation applies to apparent motion and apparent motion retrodiction results discussed earlier. As for apparent motion, the nervous system apparently has a model of how the world works that indicates that single moving stimuli are more likely than two distinct salient and quickly extinguished stimuli in very close spatial and temporal proximity. Because of this, when the stimulus conditions are right, the trajectory estimate produced at the time of the second flash is that a single moving stimulus was noisily observed by the sense organs, rather than two distinct stationary stimuli in close spatial and temporal proximity accurately observed by the sense organs. When this trajectory estimate is produced, one of the aspects of this production is that the stimulus was just at the intervening position. The percept to the effect that the stimulus was at location B was produced *after* the stimulus was observed at location C. Because of this, if time were being used to represent time, then the system would then be representing the stimulus as being at A, then C, then B. But the trajectory estimation scheme does not use time to represent time. What the system decides, at the time of the flash at C, is that the object *was* at B before its current location at C.

Now to apparent motion retrodiction. Suppose that the first pair of dots flashes at $t = 1$, the second pair at $t = 2$ and the third pair at $t = 3$. We know that if only two pairs of dots flash in the diamond-patterned bi-stable quartet, there is a 50% chance that it will be perceived as clockwise motion, and a 50% chance as counterclockwise. If, however, the third pair flashes in a location such that clockwise motion would be rectilinear, then there is a greater than 50% chance that the first two pairs are seen as clockwise motion. Again, how much greater than 50% depends on details such as the spatial and temporal distance between the dots, and other factors; see Williams *et al* (2005) for details. This entails that there are trials such that the motion appears to be rectilinear with a clockwise motion of the first pair (the bi-stable quartet), but had the third pair not flashed, the motion would have been seen as counterclockwise. What this means is that in at least some instances, at $t = 3$, the trajectory estimator retroactively modifies its prior estimate to the effect that there was counterclockwise motion and instead represents the prior motion that just occurred as being clockwise.

3.3. Representational momentum

While I have formulated the trajectory estimation model in such a way that part of the trajectory estimate at any time is a prediction of future stages of the evolution of the process' state, the examples I have used so far have not required this. A system that maintained trajectory estimates only up to and including the present time, using smoothing and filtering but no prediction, would be sufficient to account for apparent motion and the cutaneous rabbit. Indeed, as I will discuss in section 4, a mere fixed-lag smoother would be able to explain them, so long as the system was always probed for its representation after the lag had expired. I want now to motivate the prediction end of the trajectory estimation scheme.

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 and 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 is shown in figure 4.

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

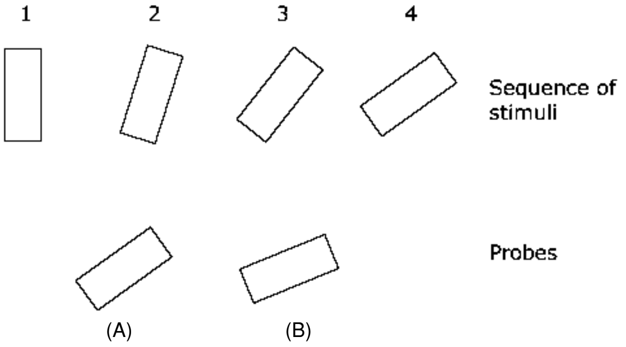

Figure 4. Representational momentum. A sequence of stimuli is shown to subjects, such as a moving ball or a rotating rectangle. The sequence is ended by a masking stimulus. Subjects are then shown two probe stimuli, such as two different end locations for the rectilinear motion, or rectangles oriented 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 and Hubbard (2002).

I will make one observation before moving on to the next section. Kalman filters are a very common way to implement forward models in control schemes. One characteristic of Kalman filters is the use of the sensor noise and process disturbance covariation matrices to determine the Kalman gain. If the perceptual systems are like Kalman filters in this regard, then this might lead to some predictions concerning the sorts of illusions I have discussed. In particular, if the perceptual system can, so to speak, adjust the gain as conditions dictate, then these illusions should be enhanced in conditions where subjects are acclimated to relatively high sensor noise, and reduced in cases where they are acclimated to relatively high process noise. Acclimating to high sensor noise is, in Kalman gain terms, trusting the internal model's predictions more than the observed signal. Since the illusions are cases where a model-based prediction is over-ruling an observed signal, these illusions should be enhanced in these conditions. For example, superimposing white noise over the computer display should increase the representational momentum effect, or the strength of the apparent motion retrodiction effect, at least if the perceptual system is like a Kalman filter in this regard.

4. Comparison with other models

It will prove helpful to compare the trajectory estimation model with two other proposals that have been applied to the case of temporal illusions: Dennett and Kinsbourne's 'multiple drafts' model (Dennett and Kinsbourne 1992), and Rao, Eagleman and Sejnowski's *smoothing* model (Rao *et al* 2001).

Using phenomena such as the cutaneous rabbit and apparent motion as motivation, Dennett and Kinsbourne argue against a conviction that they claim is nearly ubiquitous, but seldom recognized as a substantive assumption. It concerns the temporality of perception, and in particular the idea that, at least for perceptual processes, time is represented by time. The assumption, which certainly holds at large time scales, is that my perceptual belief that event A preceded event B is justified by, and perhaps amounts to no more than, the fact that I perceived A before I perceived B. Here, the temporal features of the represented content (that event A preceded event B) is a function of temporal features of the representing mechanisms—the neural states that embodied the perception of event A occurred before the neural events that embodied the perception of event B. At large time scales this is relatively unproblematic. Surely my belief that the hour hand's pointing at '1' long preceded its pointing at '4' is at least in major part explained by the fact that I perceived it pointing at '1' long before I perceived it pointing at '4'.

But at shorter time scales, the assumption that this principle still holds is not so obvious, and leads to difficulties. During apparent motion, I perceive the moving dot as being at the half-way point before it is at the end point of the motion. Otherwise it would not seem like it was moving from the initial point to the end point, but would look like it transported from the initial to the end point and then bounced back half way. Is this to be explained by my perceiving the dot at the half-way point before I perceive it at the end point? Before I perceive the end point of the motion, how can my perceptual system know whether it should interpolate motion at all, let alone in what direction? And if it does not know this, how can it correctly 'fill in' the intermediate stages of the motion?

According to Dennett and Kinsbourne's multiple drafts model, the perceptual system has the capacity to produce more than one 'draft' or judgment concerning what has been observed, and a draft produced at one time may be rewritten at a later time if additional information comes in to suggest that the initial judgment was mistaken. But such rewriting does not, on this model, require a representation of updated versions of the sensory states that led to the older, outdated draft. For example, after three taps have been sensed on the wrist, a draft containing the perceptual judgment to the effect that three taps have just been felt on the wrist is produced. Later, after some further taps have been delivered near the elbow and shoulder, a new draft is written to the effect that there *was* an evenly spaced sequence of taps beginning at the wrist and moving upward. Crucially, though, this new draft, which embodies the perceptual judgment as of the time of its production, does not require the system to go back in time and re-tool the sensory episodes, producing 'fake' sensed taps between the wrist and elbow. If asked, the subject will report a sequence of evenly spaced taps. Having a perceptual 'draft' to the effect that there was such an evenly spaced sequence just is, on this model, what it is for it to seem to the subject as though she has just experienced an evenly spaced sequence of taps.

It should be clear that the trajectory estimation model is entirely compatible with the multiple drafts model. Both claim that the representation produced at any time by the perceptual system can be overridden later on as new information is made available. A phase of a trajectory estimate made at one time, say $\hat{p}_{5/5}$, produced at $t = 5$ about what is happening at $t = 5$, may be overridden by an incompatible assessment, say $\hat{p}_{5/6}$, produced at $t = 6$, concerning what *was* happening at $t = 5$. Just as on the multiple drafts model, there is no requirement that the bare observed signals that lead to the various trajectory estimates themselves need to be represented in a different form that is compatible with the new judgment.

The differences between the trajectory estimation model and the multiple drafts model are largely differences of focus and explicitness. The trajectory estimation model is not formulated from copyediting metaphors, but on apparatus that applies internal forward models for purposes of control, signal processing and estimation. It is thus better suited for the construction of concrete simulations. Second, because of the tools used to formulate the model, it is capable, in principle at least, of being more easily integrated into current work in motor control and perceptual processing that use similar conceptual tools. Third, the trajectory estimation model explicitly highlights something only implicit in the multiple drafts model—the crucial role played by knowledge of how the process, in the cases discussed as examples this is the environment and various sorts of things in it, can be expected to behave.

Now consider the smoothing model of Rao, Eagleman and Sejnowski (Rao *et al* 2001). Rao *et al* take as their explanandum not any of the phenomena mentioned so far, but a related illusion, the *flash-lag effect*. The details of the flash-lag effect are not important for current purposes. What is relevant is that in order to shed light on the psychometric details of this effect, Rao *et al* propose that the visual system implements a fixed-lag smoother. The comparison of some modeling results and actual data from human subjects leads Rao *et al* to the conclusion that in the case of the visual system, the lag is around 80–100. As Rao *et al* put it, "The smoothing model demonstrates how the visual system may enhance perceptual accuracy by relying not only on data from the past but also on data collected from the immediate future of an event."

The fixed-lag smoothing hypothesis can also address the temporal illusions I have here discussed. If the perceptual system waits 100 ms before committing to an interpretation of what has happened, then it has access to the fact that a flash was sensed at location C before it has to start filling in interpolated motion from location A.

There is an obvious similarity between the smoothing model and the trajectory estimation model. On the trajectory estimation model, the lagging edge of the trajectory estimate is produced by a fixed-lag smoother. This entails that once the lag time has elapsed, the two models will yield identical estimates. The trajectory estimation model thus inherits all confirmatory data that speak in favor of the smoothing model, at least all data that concern estimates constructed by the systems after the lag time has passed, and the psychometric and modeling data presented by Rao *et al* are all of this nature.

This significant similarity aside, there are a number of considerable advantages of the trajectory estimation model over the fixed-lag smoother model. First, the smoothing model posits a costly *delay* in perceptual processing. The delay is costly in two respects. First, there is a computational cost involved in smoothing over filtering. More data are involved, and arriving at the smoothed estimate involves more processing than a merely filtered estimate, since the filtered estimate must be computed in route to computing a smoothed estimate. Second, there is a cost to the organism in terms of behavioral timeliness if percepts potentially crucial for action are delayed. Rao *et al* are aware of the fact that such a delay is costly, and conjecture that the magnitude of the lag (about 80–100 ms) might represent an optimal tradeoff between the cost of delay, and the benefit of more accurate percepts.

The trajectory estimation model need make no such conjecture, because on this model nothing is delayed, and hence there is no cost of behavioral timeliness. Not only are percepts not delayed, but indeed they are *anticipated*. It remains true that the specific state that is anticipated or estimated might be revised as more data come in. But openness to revision for, say, 100 ms, should not be confused with delaying all interpretation until 100 ms has passed. The functional import of the lag in the trajectory estimation model is that it is the deadline for incoming information to be able to influence the ongoing trajectory estimate. But the trajectory estimate is up and running the entire time. It seems plausible to suppose that two of the most important functions of nervous systems as they evolved are anticipation and accuracy. The smoothing model sets these desiderata up at cross purposes, sacrificing time for accuracy. The trajectory estimation model embraces them both simultaneously.

So there is no behavioral timeliness cost. As for the computational cost, the trajectory estimation model is costlier still than the smoothing model. It not only requires filtering and smoothing, but also prediction. But this cost is presumably offset by the lack of an additional cost from perceptual delay and the benefit of perceptual anticipation. It is not clear how to assess this balance of costs and benefits, so I will end the speculation here.

A related advantage is that the smoothing model seems to imply that we are under an illusion of control in some contexts. It *seems* to us as though we do certain things on the basis of what we perceive. But for anything that gets executed in less than 100 ms, this must be mere illusion if the smoothing model is correct, since on this model we are not conscious of visual percepts until after 100 ms or so has elapsed. This by itself is no *major* disadvantage, since how things seem to us in such contexts can hardly be taken to be definitive. That we are under such illusions is a very live possibility. However, though not a major disadvantage, it would seem that if an account that has equal or better explanatory potential is available, and lacks this counter-intuitive result, then this ought to count in favor of the alternate proposal. Surely being counter-intuitive is not, by itself, a theoretical desideratum! The trajectory estimation model is not forced to take a stand on the issue of the veracity of the sense of conscious control. Since percepts are available in real time and indeed even anticipated, the possibility of online conscious control based on perceptual representations is not ruled out. We are not, as David Eagleman has provocatively put it, 'living in the past' by 80 ms, or any other amount of time.

I will remark now on one *prima facie* advantage that the smoothing model has over the trajectory estimation model. The proponent of the smoothing model has an answer to the question why the lag has the particular magnitude it has. That speculative answer, as I mentioned above, is that a lag on the order of 100 ms or so might represent something like an optimal tradeoff between the cost of delaying perceptual processing and the benefit of more accurate percepts. Since on the trajectory estimation model there are no delayed percepts, there is no cost, except computational, to extending the lagging edge of the estimated trajectory arbitrarily. Is there any reason, even an equally speculative one, as to why the temporal interval spanned by the trajectory estimator might have lagging and leading edges on the order of 100 ms into the past and future respectively?

First to the lagging edge. The point of smoothing is to allow information collected after some event to help aid the interpretation of the observation of that event. If there were some delay period such that information collected within that period could be expected to have a significant impact on the smoothed estimate, but beyond that delay it was unlikely that further information would have a significant impact, this would motivate setting that delay period as the lag. In biological organisms there is such a time period—the longest sensory transmission delay. All the sensory information is carried to the central nervous system via neural signals, which are relatively slow. Proprioceptive information from the feet is probably the most significantly delayed information, and while the exact magnitude of this delay is subject to debate, it appears to be on the order of 100 ms. This means that the perceptual processors can expect, as a matter of course, that upon constructing a state estimate of the body at time t, information will continue to arrive up to $t + 100$ ms or so that will be directly relevant to the accuracy of that estimate. Beyond that delay, all of the relevant sensory information will typically be on hand. So the longest typical sensory delay sets a natural point to anchor the lag.

Now to the leading predictive edge. Predictions are only useful to the extent that they are accurate. Is there some prescience boundary such that up to that boundary certain kinds of predictions are much more reliably accurate than predictions that transcend that boundary? Again, neural conditions velocities provide the key. It is plausible to suppose that efference copies of the organism's own motor commands, together with an internal model of the body, provide the material for very reliable estimates of the state of the body upon execution of that motor command. But those motor commands take time to propagate to the musculature and have their effects. Again, this delay is on the order of 100 ms or so (for voluntary actions; reflexes are faster, of course). This means that at time t, when the system has an estimate of the body's current state and has an efference copy of the currently issued command, it will be in a position to produce a very reliable estimate of the state of the body at $t + 100$ ms or so. Recall that it was pointed out in section 2 that predictions for future states of a driven Gauss–Markov process require knowledge of the driving force that will be in effect at the time for which the prediction is attempted. These considerations are extremely speculative, of course. The point is merely to illustrate that the trajectory estimation model is not without resources for addressing the question of the apparent magnitude of the temporal interval of the estimated trajectory.

5. General discussion

In the introduction I remarked on the fact that although proponents of internal models are at theoretical odds with adherents to the embedded cognition camp, in the case of time they are in agreement. Time is typically used to represent time in internal modeling schemes. I have tried to show that it is neither necessary nor advisable for the proponent of internal models to acquiesce on this matter. It is not necessary because the possibility of internal models of trajectories over temporal intervals, as opposed to internal models of states at a time, has the capacity to represent time by means other than time. And it is not advisable because the existence of temporal illusions appears to be good *prima facie* evidence to the effect that time is explicitly represented by the perceptual system. Something has to be *represented* in order for it to be *mis*represented. There are several aspects of this capacity for temporal representation worth mentioning.

The first point has to do with the paradoxical point that tools from filtering and optimal estimation are being used to explain the existence of *illusions*! This is true not only of the trajectory estimation model, but Rao *et al*'s smoothing model as well. Indeed, the paradox is stronger on the smoothing model, since it is precisely the ability to *improve* estimates that smoothing is motivated, but the phenomenon it is addressing is our blatant mis-estimation of the location of a stimulus. What this point highlights is the role played in these schemes by the model of the process that is used by the internal model, whether it be for filtering, smoothing, prediction or trajectory estimation. The process model embodies expectations, presumably learned through observation, in the form of the function that describes how the process—the body and the environment—evolves over time. It is by exploiting this knowledge that such systems are able to produce estimates that are able to reduce one or another sort of expected error. Illusions are cases where the environment is comporting itself, sometimes with an experimentalist's aid, in a statistically irregular way with the result that the expectation embodied in the process model leads the estimation process astray. The paradox is merely apparent.

The second point has to do with the nature of the temporal interval that is represented. Since my goal was merely to introduce the basic concepts rather than to develop a specific model, I left most details out. But the empirical data appear to suggest that the interval spanned by the trajectory estimator that handles human perception is roughly on the order of 200 ms or so, from about 100 ms in the past to about 100 ms in the future. For the past direction, the Rao *et al* model, which is here being interpreted as exemplifying the special case of the lagging edge of the estimated trajectory, suggested that the lag was about 100 ms. This is consistent with the Williams *et al* apparent motion retrodiction result, which found that the effect was only found if the last pair of flashing dots was presented within 100 ms. As for the leading predictive edge of the estimated trajectory, representational momentum would appear to be the obvious phenomenon to examine. The problem is that the amount of shift is somewhat variable. Nevertheless, it seems that a fraction of a second, roughly on the order of 100 ms or so, is in the ballpark.

The third point is that if the trajectory estimate that is produced at any given time is what one perceives at any given time, then the trajectory estimation model is an instance of the *specious present* doctrine. This is a doctrine, given currency in psychology by William James in his highly influential *Principles of Psychology* (James 1890). The doctrine maintains that what we are aware of at any instant is not a corresponding instant of the perceived environment, but a span of time. As James put it:

> In short, the practically cognized present is no knife-edge, but a saddle-back, with a certain breadth of its own on which we sit perched, and from which we look in two directions into time. The unit of composition of our perception of time is a *duration*, with a bow and a stern, as it were—a rearward- and a forward-looking end. It is only as parts of this *duration-block* that the relation of *succession* of one end to the other is perceived. (James 1890, pp 609–10)

The doctrine has been appealed to in order to explain our capacity to perceive motion, for example the motion of a second hand. While we can come to judge that an hour hand is moving by comparing its position as we now perceive it to its position as we remember it from some time ago, it seems as though we can simply directly see the motion of a second hand. But motion takes time, and so the content of our visual experience appears to span at least a duration sufficient for us to notice the second hand's movement. The specious present doctrine is controversial, and my purpose is not to enter into that debate here, but merely point out that if the trajectory estimation model is right, then at least one version of the doctrine will be vindicated.

The final issue I wish to raise concerns the role of neural 'timing' mechanisms. When the issue of how time is represented by nervous systems is raised, quite often it is assumed that the existence of time-keeping mechanisms is what is being asked after, something like a neural clock (Hazeltine *et al* 1997, Nenadic *et al* 2003). There are two distinct uses to which a nervous system might put a time-keeping mechanism. First, the mechanism could simply help to guarantee that the timing of behavior is appropriate—everything from timing a hand clap with the beat of music to tracking time in order to use the angle of the sun correctly as a navigational marker (Froy *et al* 2003).

A second use of time-keeping mechanisms is more relevant for the issue of internal models. When the internal process model evolves its state from one time step to the next, the function that affects this mapping should be calibrated such that it will evolve the process model's state in a way that mirrors, as closely as possible, the evolution of the state of the real process *over that same amount of time*. That is, if each update of the process model's state estimate takes 20 ms, then the function that is used to update the state of the process model ought to change the process model's state to mirror the change that the actual process undergoes in 20 ms. Obviously, if every 20 ms the internal model's state is updated to reflect a change that the real process would undergo in 40 or 100 ms, the *a priori* estimates will not be very accurate. This time-tracking capacity is a matter of calibrating two separate aspects of the internal model: (i) the function that maps state estimates onto successor state estimates, and (ii) the timing of the production of successor state estimates. In neural implementations of process models, the second element here is most likely governed by some combination of the intrinsic dynamic properties of the neural structures that implement the model and timing mechanisms, such as oscillators that cycle at more or less regular intervals. Exactly similar points hold for *trajectory* estimation schemes of course.

6. Conclusion

My goal in this paper has been modest. I have not tried to produce a specific detailed trajectory estimator model of perception. Such a task would involve some specific choices as to what the span of the temporal interval is, the specific nature of the process models implemented by the system, and choices as to whether to model it as a generalized Kalman filter/smoother/predictor, some sort of optimal tuned filter, or perhaps one that is suboptimal in one or more ways. A few speculations on these specifics have emerged. The magnitude of the interval can probably be ascertained empirically via phenomena such as representational momentum, which perhaps provide a clue as to how far in the future predictions are produced, and considerations such as those mentioned by Rao *et al* that appear to single out some maximum lag for the smoothed end of the trajectory estimate. It seems unlikely that the nervous system is strictly optimal, especially since it most likely does not have access to the process disturbance and sensor noise covariation matrices, as required by the vanilla Kalman filter. And it also seems unlikely that the nervous system has an implementation of an accurate and complete system identification of the environment, as the Kalman filter requires, and hence is more likely to implement a variety of tuned filters, tuned to various sorts of motion and force-dynamic interactions, for example.

Working out these details is surely required for determining whether some sort of trajectory estimation model accurately describes some aspect of the nervous system's perceptual processing mechanism. But my goal in this paper has merely been to describe, in general and schematic terms, what a trajectory estimation model would look like, and to indicate, again in rough terms, how such a model might be able to shed considerable theoretical light on a range of puzzling phenomena. I have also tried to indicate how such an approach compares favorably to two other models that have been proposed to address the same phenomena. Hopefully this schematic framework will be seen as promising enough to merit further investigation and clarification by those researchers whose work touches on the temporal features of perception.

References

Ashby W R 1952 *Design for a Brain: The Origin of Adaptive Behavior* (London: Chapman & Hall)

Brooks R 1991 Intelligence without representation *Artif. Intell.* **47** 139–59

Bryson A and Ho Y-C 1969 *Applied Optimal Control; Optimization, Estimation, and Control* (Waltham, MA: Blaisdell)

Chomsky N 1959 Review of Skinner's *Verbal Behavior. Language* **35** 26–58

Dennett D C and Kinsbourne M 1992 Time and the observer: the where and when of consciousness in the brain *Behav. Brain Sci.* **15** 183–247

Freyd J J and Finke R A 1984 Representational momentum *J. Exp. Psychol.* **10** 126–32

Froy O, Gotter A L, Casselman A L and Reppert S M 2003 Illuminating the circadian clock in monarch butterfly migration *Science* **23** 1303–5

Geldard F A and Sherrick C E 1972 The cutaneous 'rabbit': a perceptual illusion. *Science* **178** 178–9

Gengrelli J A 1948 Apparent movement in relation to homogeneous and heterogeneous stimulation of the cerebral hemispheres *J. Exp. Psychol.* **38** 592–9

Grush R 2003 In defense of some 'Cartesian' assumptions concerning the brain and its operation *Biol. Phil.* **18** 53–93

Grush R 2004 The emulation theory of representation: motor control, imagery, and perception *Behav. Brain Sci.* **27** 377–442

Haugeland J 1995 *Mind embedded and embodied Mind and Cognition: Philosophical Perspectives on Cognitive Science and Artificial Intelligence (Acta Philosophica Fennica* vol 58*)* ed L Haaparanta and S Heinäma pp 233–67

Hazeltine E, Helmuth L L and Ivry R 1997 Neural mechanisms of timing *Trends Cogn. Sci.* **1** 163–9

Ito M 1970 Neurophysiological aspects of the cerebellar motor control system *Int. J. Neurol.* **7** 162–76

Ito M 1984 *The Cerebellum and Neural Control* (New York: Raven Press)

James W 1890 *The Principles of Psychology* (New York: H Holt and Company)

Kerzel D 2002 A matter of design: no representational momentum without predictability. *Vis. Cogn.* **9** 66–80

Kolers P A 1972 *Aspects of Motion Perception* (London: Pergamon)

MacKay D M 1958 Perceptual stability of a stroboscopically lit visual field containing self-luminous objects *Nature* **181** 507–8

Miall R C, Weir D J, Wolpert D M and Stein J F 1993 Is the cerebellum a Smith predictor? *J. Motor Behav.* **25** 203–16

Nenadic I, Gaser C, Volz H-P, Rammsayer T, Häger F and Sauer H 2003 Processing of temporal information and the basal ganglia: new evidence from fMRI. *Exp. Brain Res.* **148** 238–46

Ramachandran V S and Anstis S A 1983 Extrapolation of motion path in human visual perception *Vis. Res.* **23** 83–85

Rao R, Eagleman D and Sejnowski T 2001 Optimal smoothing in visual motion perception *Neural Comput.* **13** 1243–53

Skinner B F 1935 Two types of conditioned reflex and a pseudo type *J. Gen. Psychol.* **12** 66–77

Thornton I M and Hubbard T L 2002 Representational momentum: New findings, new directions *Vis. Cogn.* **9** 1–7

Tolman E C 1948 Cognitive maps in rats and men *Psychol. Rev.* **55** 189–208

Verfaillie K and Daems A 2002 Representing and anticipating human actions in vision *Vis. Cogn.* **9** 217–32

Watson J 1913 Psychology as the behaviorist views it *Psychol. Rev.* **20** 158–77

Wiener N 1948 *Cybernetics; or, Control and Communication in the Animal and the Machine* (New York: Wiley)

Williams L E, Hubbard E M and Ramachandran V S 2005 Retrodiction in apparent motion (submitted)