Brain time and phenomenological time

Rick Grush
Department of Philosophy, UC San Diego
Word count: 14,359
Version 3.0
Revision: 05.08.2003

> As 'tis from the disposition of visible and tangible objects we receive the idea of space, so from the succession of ideas and impressions we form the idea of time...
>
> - David Hume, *Treatise of Human Nature*

> ... space and time ... are therefore pure intuitions that lie *a priori* at the basis of the empirical. ... they are mere forms of our sensibility, which must precede all empirical intuition, or perception of actual objects ...
>
> - Immanuel Kant, *Prolegomena to Any Future Metaphysics*

> ... there are cases in which on the basis of a temporally extended content of consciousness a unitary apprehension takes place which is spread out over a temporal interval (the so-called specious present). ... That several successive tones yield a melody is possible only in this way, that the succession of psychical processes are united "forthwith" in a common structure.
>
> - Edmund Husserl, *Phenomenology of Inner Time Consciousness*[1]

1. Introduction

The topic of this paper is temporal representation. More specifically, I intend to provide a theory of what it is that our *brains do* (at the sub-personal level) such that *we experience* (at the personal level) time in the way that we do. A few words about both sides of this relation are in order.

As far as the brain goes, I will actually be making little substantive contact with neurophysiology. The main thrust of my strategy on the brain side is to articulate an

---

[1] This particular passage is one in which Husserl is articulating, with approval, a view he attributes Stern (1897).

information-processing structure that accounts for various behavioral and phenomenological facts. The neurophysiological hypothesis is that the brain implements this information-processing structure. The amount of neurophysiology won't be zero, but at this stage of the game, our understanding of the brain's capacities for temporal representation are incredibly slim. The experimental side of neurophysiology is in need of some theoretical speculations to help it get going in earnest.

As far as our personal-level experience goes, there are only a few central aspects that I will be addressing. It will help to mention some of the aspects I will *not* be addressing. I won't be addressing memory, including how it is that our memories come to us with the conviction that they concern events that happened long ago. Nor will I be concerned with what might be called objective temporal representation. My belief that Kant's *Critique of Pure Reason* was published less than 100 years after Locke's *Essay Concerning Human Understanding* does depend on my ability to represent objective temporal relations and objective units of time, such as years. But the capacities involved in such temporal representation are not my concern.

Rather, I am directly interested in what I will call *behavioral time*. This is the time that is manifest in our immediate perceptual and behavioral goings on, and in terms of which these goings on unfold. I will expand on this shortly, but first an analogy with spatial representation may prove helpful. In philosophical, psychological, and cognitive neuroscientific circles it is common to distinguish allocentric and egocentric spatial representation. The contrast is between an ability to represent things that in no way depends on my own whereabouts, and the representation of things in relation to myself. So my ability to represent the Arc de Triomphe as being between the Obelisk and La Grande Arche de la Defense in no way depends on my own location in space. Whereas my belief that there is a pitcher's mound 90 feet west of me relies on my own location as a sort of reference point.

But there are *two* senses in which a spatial representation can be non-objective. In the sort of case that I have called egocentric, the location of objects is represented in relation to oneself, rather than as being at some objectively specifiable spot in an objective framework. But the units and axes of such a specification might yet be objective. My belief that the pitcher's mound is 90 feet west of me is not objective in that it makes reference to my own location as a reference point, but the axes and units employed in this specification *are* objective.

I use the expression *behavioral space* for a kind of spatial representation in which not only the reference point, but also axes and magnitudes that define the space are non-objective. My representation of the coffee cup as being *right there* when I see it and reach out for it specifies its location relative to me, in a space whose dimensions are spanned by axial asymmetries of up/down, left/right, and front/back -- axial asymmetries whose content derives from my own behavioral capacities. And the magnitudes involved -- the difference between the cup that is *right there* and the sugar bowl that is *over there* -- are also imbued with content via their connections to my own behavior. I may have no clear idea how far the coffee cup is from me in inches and feet, but I have a very precise representation of its distance specified in behavioral terms, as is evident from the fact that I can accurately grasp it.

Back to time. Analogues of allocentric/objective, egocentric and behavioral space are readily specifiable in the temporal domain. Allocentric/objective temporal representation is exploited by my belief that Kant's masterwork was published 90 years after Locke's; and also in my belief that the numeral '4' always appears in the seconds position of my watch one second after the numeral '3' appears there. Egocentric temporal representation is involved in my belief that Kant's first *Critique* was published 222 years ago (i.e. back from *now*); and it is also manifested when I see the numeral '3' appear in the seconds spot of my stopwatch and I come to believe that the numeral '4' will appear one second from *now*. Egocentric temporal representation uses my current time, *now*, as a temporal reference point much like egocentric spatial representation uses my current location, *here*, as a spatial reference point. But the *behavioral* time specifies the temporal dimension and magnitudes not in terms of such objective units, but in terms of behavioral capacities. When I move to intercept and hit a racquetball that is moving quickly through the court, I may have no accurate idea, in terms of seconds or milliseconds, of how far in the future the point of impact between my racquet and the ball will be. But I am nevertheless quite aware in behavioral terms. My movements and planning reveal an exquisite sensitivity to the temporal features of the event that will unfold. A more common example might be moving one's hand down to catch a pencil that has just rolled off the edge of a table. One's attention is palpably focused on a spatio-temporal point -- *just there* and *just then* (a foot or so from the torso and a few hundred millliseconds in the future, though the units are not in terms of feet or milliseconds, but are behaviorally defined) -- at which the hand will contact the pencil.[2]

---

[2] Further analogies between on the one hand allocentric and behavioral space and on the other hand allocentric and behavioral time suggest themselves. I will confine my own phenomenological introspections to a footnote: Sudden coordiantions of allocentric and

It should be clear that the distinction between these three kinds of representation is not a matter of magnitude. I can represent quite large and quite small spatial and temporal intervals allocentrically and egocentrically, especially since the units in both cases are objective units. My belief that Kant's masterwork was published 222 years ago, and my belief that the numeral '4' will appear in the seconds position of my watch are examples. In the case of behavioral space and time large and small magnitudes can be represented, though because the units derive from perception and behavior, discriminatory abilities are best within a certain behaviorally relevant range both spatially and temporally and fade off as we exceed the limits of what typically becomes behaviorally manifested, both with very large spatial and temporal magnitudes, as well as very small spatial and temporal magnitudes.

Not only will I be limiting myself to behavioral time in this paper, I will also be focusing my attention on a very limited but special region of behavioral time, the *behavioral now*. There is reason to believe that a small temporal interval, spanning a few hundred milliseconds into the past and future, has a special kind of status both psychologically, philosophically and neurophysiologically. The locus classicus in psychology is William James' 'specious present', though the behavioral now that I am interested in is much shorter. In the philosophical tradition Edmund Husserl was the first to examine time consciousness in detail. He posited a structure of primary protentions and retentions as brief temporal regions in which the mind is actively synthesizing temporally conditioned representations: anticipations of what will be unfolding in the immediate future as well as retentions of what has just happened. For Husserl, protentions and retentions are substantive features of what is being experienced *now*. Daniel Dennett has pointed out a number of pitfalls in our understanding of the representation of time, and has argued that within some small interval[3] the brain is an active interpreter of temporal experience, and not a passive 'Cartesian theatre'. In cognitive

---

behavioral spatial representations are phenomenologically quite strong. Scandinavian Airlines has the habbit of projecting a map of the aircraft's progress during the flight. On more than one occassion, as I saw the small outline of the aircraft pass by soome interesting terrestrial feature, such as Iceland, I would make the realization "Oh, Iceland is now *down there*!", where the 'down there' was identified in my behavioral space: a sudden feeling of knowing where things are. I have had similar experiences in the case of time. For example, times during which though I have known in egocentric terms that a paper had to be submitted within 15 days, or that I had to give a presentation on Monday, and then I suddenly made a coordination of this with my behavioral time so that I could, so to speak, *feel* how close the event was. This iis often accompanied by a sudden adrenaline rush as I realize in more behaviorally salient way just how little (behavioral) time is left.

[3] Dennett is not as clear as one might like, but the implication seems to be is that there is an interval such that at temporal scales smaller than this interval the distinction between 'Orwellian' and 'Stallinesque' temporal interpretation breaks down. See Dennett 1992; Dennett and Kinsbourne 1992.

neuroscience, there has been a recent flurry of activity on the temporality of perception that suggests that the brain is an active temporal interpreter. For example Eagleman and Sejnowski have shown that what one perceives as having happened at *t* depends in part on what happens with the stimulus within some small interval after *t*. (Dennett and Eagleman et al. will be discussed in more detail in Section 7.)

The next five sections construct the theory. Sections 2 and 3 concern not time but space. There are two reasons for the spatial detour. The first is that the account of temporal representation is parallel in many respects to an account of spatial representation that I have articulated elsewhere, and various of its features are more perspicuous in the case of space. Section 2 details a theory of how neural states can come to carry spatial information, and Section 3 provides an account of how states that carry spatial information can come to have spatial import for the subject.

In sections 4 and 5 I develop a theory of how neural states can carry temporally conditioned information. Section 4 recaps the emulation theory of representation (Grush, 2003; to appear), which is an information processing framework according to which the core of the brain's representational machinery involves forward models, especially as used in a Kalman-filter control schemes. Section 5 extends that framework from mere filtering to a combination filter-smoother-predictor that maintains information about a represented system over some temporal interval, rather than a temporal point like the bare filter does. Section 6 then explains what is involved in such states being able not only to carry temporal information, but having temporal import for the subject.

In the final Section 7 I will return to a general discussion. First I will compare the theory presented here with some recent work in temporal perception by Eagleman and Sejnowski (Eagleman and Sejnowski 2000) and its relation to some of Dennett's views on temporal representation (Dennett 1992; Dennett and Kinsbourne 1992). Second, I make a few speculations co0ncerning why such an information processing scheme I describe might have evolved, and why it might have the specific temporal dimensions that it has. And finally I will make some remarks meant to show how the theoretical stance of this paper can be seen as part of a larger program in cognitive science that embraces a Kantian approach to core elements of the content of the mind, rather than the more typical Humean approach.

2. Spatial information

There are many kinds of spatial representation, and even when the scope is narrowed to what neuroscientists and psychologists call egocentric space (by which they mean something more like what I have described as behavioral space than egocentric space), the neural areas implicated are many and complex. Nevertheless, the posterior parietal cortex (PPC) has emerged as perhaps the most important cortical area involved in the representation of behavioral space. Damage to this area, especially the right PPC in humans, leads to a characteristic pattern of deficits that to a first approximation can be described as an impairment of the capacity to represent contralateral behavioral space. But though the behavioral deficit data suggest that this is what the PPC is doing, single cell recordings from the area do not reveal anything like a map of the behavioral space. Given that there are topographic maps of many other sorts throughout the brain, one might have expected there to be a topographic map of the behavioral space. But instead, neurons in this area appear to correspond to things like sensory stimulation at point on the retina; or the stretch of eye muscles; or head orientation. What the connection is between things like this and a representation of the behavioral space might not be obvious. But as in all things, the medium of clarity is a good theory.

The major breakthrough in understanding the PPC came in work by Zipser and Anderson. These researchers trained a connectionist network to solve a toy sensorimotor problem which was as follows: an artificial organism with one rotatable eye was to determine the location of a stimulus that projected onto its retina. The difficulty is that because the eye can rotate, simply knowing the location of stimulation on the retina does not determine a unique stimulus location. The same stimulus location on the retina might have been caused by any of a number of stimulus locations depending on the orientation of the eye. However, given the location of stimulation of the retina *and* the orientation of the eye with respect to the head, the location of the stimulus relative to the head can be determined.

With this in mind, the neural network was trained by being given two inputs: location of retinal stimulation and eye orientation. And from these it was to correctly ascertain the location of the external stimulus relative to the head. During training, the correct answer was provided in the usual backprop sort of way. After it learned to solve the problem the network could be analyzed in detail to see how it ticked -- one of the big advantages of artificial networks over their biological counterparts. The result was that the 'hidden' units (the units between the input and output units that, in many connectionist architectures, is where the interesting work happens) appeared to be acting as linear gain fields. That is, the activity of a given hidden unit would have its activity determined by two factors: i) a gaussian function of distance between the location of the stimulus' projection on the retina and some preferred spot on the retina, such that if the stimulus falls on this preferred spot the activity is highest, and the activity falls off with the stimulus' distance form that location; and ii) a linear function of preferred eye orientation -- highest if the eye is oriented towards the preferred direction (left or right) and lower as it is oriented in the opposite direction. The net result is that a given hidden unit's activity is most strong for a certain combination of retinal location of stimulation and eye orientation, and such a combination corresponds to a location in head-centered space. (The details are a bit more complicated, but these are the essentials.)

Neurophysiological studies verifies that actual PPC neurons has response profiles that fairly closely matched these, enough to support the idea that perhaps the PPC neurons were processing information in more or less the same way. And not only does the gain field hypothesis explain how sensory signals can be combined with signals concerning eye orientation to determine a location relative to the head, but the mechanism generalizes to providing information about spatial location with respect to the head or torso given the necessary postural signals concerning the orientation of the eyes with respect to the head, the head with respect to the torso.

Of course, one might ask: Why should the brain care about developing units (neurons or groups of neurons) whose activity corresponds to the location of a stimulus relative to the head? It might be thought that positing spatial representations of this sort is unnecessarily complicated. The brain can implement coordinate transformations from sensation to action without having to construct this kind of intermediary.[4] This is true, and if one restricts attention to one sensation-action episode, then it would seem like constructing this intermediary is more work than needs to be done. It would be possible to do coordinate transformations for each episode of sensorimotor activity without representing anything like the location of the stimulus relative to the head or torso. One could determine an appropriate set of arm effector commands given a certain combination of sensor platform orientations:

---

[4] Such coordinate transformations would map points in a high dimensional sensory space to points in a high dimensional motor space. The point in sensory space would carry information about, e.g. where on the retinae a stimulus is occurring, what the orientation of the eyes relative to the head, and head relative to the torso are. The points in motor space would specify patterns of activity (dynamic or kinematic, perhaps in the form of equilibrium points, whatever). And the proposal is that all of this might happen without the intermediary of units whose activity is correlated with points in head or torso centered space.

for example, if the eye orientation with respect to the head is E and the head orientation with respect to the torso is H, and the retina is being stimulated at point R, then a movement of the arm like M will get the hand to the stimulus.

But when one looks at the bigger picture and realizes that there are many sensory channels, many effectors, and coordination between the effectors is sometimes required, it becomes clear that having a common representational format is the most economical way to do things. The overall computational load is tremendously reduced by determining location relative to the head and/or torso. A torso-centered spatial representation allows for a stimulus to be placed at a location in the behavioral space regardless of whether it is seen or felt, and provides a common framework for action whether the action will involve the left hand, right hand, head and mouth, or anything else. Representation of spatial location with respect to the torso provides for a common coordinate system, so to speak, within which all perception and action can be coordinated.[5]

Next, it might be objected that while having a common representational format is useful, the idea that that format must in terms of any sort of space does not follow. In particular, it might be suggested that something along the lines of Pouget’s basis function model will do the job. I can’t go into this here, but there is no headlong conflict between the basis function model and the idea that the common representational format is spatial. The three axial asymmetries of up-down, left-right, and front-back, which span behavioral space, are not only adequate, but indeed natural bases for any kind of basis-function representation of this sort to employ.

For now it is sufficient to see how the appropriate organization of neurons and their interconnections can be made such that the activity of a single neuron, or more plausibly the pattern of activity of a group of neurons, carries information concerning spatial locations, in the way that the pattern of activity of the hidden units in the Zipser and Anderson model covaries with, and hence carries information about, the location in space of the stimulus (or how a neural representation of a given set of basis functions can carry information about stimulus location relative to the torso, as in Pouget's model). These units process information about the stimulus location as well as the relative orientation of the relevant sensory platforms (eye orientation, head orientation) in order to determine the location of the stimulus relative to the torso (or head, etc.).

---

[5] This point and the next are developed in more adequate detail in Grush (in preparation).

3. Spatial import

The last section gave a very brief explanation of how it is possible for states of neural systems to carry information about locations in space relative to the torso. But carrying information about such things cannot be the whole story of our spatial perceptual experience, for surely we can have perceptual or experiential states that carry spatial information but lack spatial import: that is, are such that we don't experience them *as* spatial.

To help me explain what I mean here I will introduce the *sonic guide*, a sensory substitution device designed to provide blind subjects with an artificial distance sense that is in some ways analogous to vision.

[Figure 1]

The sonic guide has a transmitter worn on the head with a speaker that produces a constant ultrasonic probe tone. A stereophonic microphone, also mounted on the headgear, picks up the echoes produced by the reflection of this probe tone off objects in the subject's vicinity. A processor takes the information gained through these echoes and translates it into audible sound profiles that the subject hears through stereo earphones. There are four aspects to the translation:

> 1. Echoes from a distance are translated into higher pitches than echoes from nearby. E.g. as a surface moves toward the subject, the sound it reflects will be translated into a tone which gets lower in pitch.
>
> 2. Weak echoes are translated into lower volumes. Thus as an object approaches the subject, the subject will hear a tone which increases in volume (as the object gets closer, it will ceteris paribus reflect more sound energy, resulting in a stronger echo), and gets lower in pitch (because it is getting closer, as in (1) above). An object which grows, but stays at a constant distance, will get louder, but stay at the same pitch.
>
> 3. Echoes from soft surfaces – e.g. grass and fur -- are translated into fuzzier tones, while reflections from smooth surfaces – e.g. glass and concrete -- are translated into purer tones. This allows subjects to distinguish the lawn from the sidewalk, for example.

> 4. The left-right position of the reflecting surface is translated into different arrival times of the translated sound at each ear. (Note that it is not required that this coding exploit the same inter-aural differences which code direction in normal subjects. In fact, if the differences are exaggerated substantially by the guide, then one would expect a better ability to judge angle than we typically have.)

As John Heil (1987) describes it, the

> …sonic guide taps a wealth of auditory information ordinarily unavailable to human beings, information that overlaps in interesting ways with that afforded by vision. Spatial relationships, motions, shapes, and sizes of objects at a distance from the observer are detectable, in the usual case, only visually. The sonic guide provides a systematic and reliable means of hearing such things.

The first lesson I want to draw with the aid of the sonic guide has to do with the distinction between spatial information and spatial import. If you or I were to don the guide, we would be in receipt of a great deal of information about the spatial characteristics of and relations between objects and surfaces in our vicinity. For example, a bowling ball on the ground a few meters in front of us would cause us to hear Middle C at 35dB, say. Our experience of this sound carries spatial information, in that the features of this sound covary in the required ways with the location of the ball. Not only would the states we experience (describable as say Middle C at 35dB) carry such information, but neural states (describable in terms of firing frequencies and patterns in some pool of neurons) would also of course be carrying exactly the same information. The pattern of neurons on the auditory pathways and cortical areas that fire when and only when Middle C at 35dB is heard will be a neural pattern that carries information to the effect that a roughly bowling-ball sized object is 2 meters ahead on the ground. So at both the personal experiential level and sub-personal neural levels there are states that carry spatial information about the location of the ball in egocentric space. And yet for all, that you or I would not experience this sound as having any significant spatial import. It would be merely a particular pattern of sound heard through earphones. This should make the distinction between carrying spatial information and having spatial import for the subject clear enough.

Nevertheless, it is possible for someone who is sufficiently practiced with the guide to be such that through it they do enjoy perceptual states with spatial import. I will suppose what seems to be plausible, that subjects who have been using the device for a while and are competent with it are actually *perceiving* the objects in their environment directly, rather than *reasoning out* what the environment must be like on the basis pitches and volumes. This seems to be accepted by Heil who, in discussing the sonic guide and another sensory substitution device, the TVSS, notes:

> Devices like the sonic guide and the TVSS prove useful only after the sensations they produce become transparent. In employing a TVSS, for instance, one ceases to be aware of vibrations on one's skin and becomes aware, rather, of objects and events scanned by the camera. Similarly, successful use of the sonic guide requires one to hear things and goings-on rather than the echoes produced by the device. ...[children] less than about 13 months... do this quite naturally, while older children do so only with difficulty.

The question is, what is the difference between you and I, on the one hand, and the blind subject practiced with the guide and for whom its deliverances have spatial import, on the other? The quick and surely correct answer is that she is used to the guide, and we are not. But what does this mean?

I have elsewhere articulated and defended a theory of spatial content that I call the *skill theory*, and which was first expressed in more or less the form I defend it by Gareth Evans.[6] The skill theory maintains that a sensory episode becomes imbued with spatial content for an organism if that sensory episode disposes or enables the organism to behave in certain spatially significant ways. To hear a sound as coming from the left *just is* for the sound to dispose me to turn to the left in order to orient towards it, or to run to the right in order to distance myself from it (and similarly for an indefinite number of other possible behaviors). As Evans put it:

> The subject hears the sound as coming from such and such a position, but how is this position to be specified? We envisage specifications like this: he hears the sound *up*, or *down*, *to the right* or *to the left*, *in front* or *behind*, or *over there*. It is clear that these

---

[6] For Evans' view, see 'Understanding demonstratives', chapter 6 of *The Varieties of Reference*, and 'Molyneux's question'. For my own development of this view, see Grush 1998, 2000.

> terms are *egocentric* terms: they involve the specification of the position of the sound in relation to the observer's own body. But these egocentric terms derive their meaning from their (complicated) connections with the actions of the subject ... (MQ: 384)
>
> Auditory input, or rather the complex property of auditory input which codes the direction of the sound, acquires a spatial *content* for an organism by being linked with behavioral output ... (MQ: 385)
>
> ... we must say that having the perceptual information at least partly consists in being disposed to do various things ... (MQ: 383)

What does this mean in terms of neural infrastructure? The Zipser and Anderson model implicitly involves an action element in that the network is trained to make correct assessments of stimulus location given only retinal location and eye orientation. In biological practice such correct assessments would be manifest only via some behavior or other: trying to grasp the stimulus or foveate on it, for instance. In more sophisticated models, such as Pouget's basis function model, the basis functions that extract the spatial information from sensory signals are linked to behavior via sets of coefficients that, together with the perceptually determined basis functions, poise the organism to execute any of a range of spatial behaviors -- or as Evans put it, dispose the organism to do various things. (I discuss all of this in much more detail in Grush 1998, 2000).

At the sub-personal level, appropriate experience, perhaps in conjunction with favorable conditions such as starting at an early enough age, effects a coordination between the relevant discriminable sensory states and spatial behavior via the right sort of connections, presumably in the PPC, between those neural pools whose job is to put sensory information into a format capable of guiding behavior by stabilizing it with respect to a coordinate frame centered on the torso (in the central case), and those behaviors themselves. This is what the gain fields and basis functions do. The manifestation of this sub-personal machinery at the personal level is the experiencing of the stimulus as being located somewhere in the behavioral space. Thus, though we all sensorily discriminate, via hearing, Middle C at 35dB, the blind subject's auditory cortex mainlines this channel of input to the PPC in a way that extracts spatial information from it together with other relevant sensory signals such as proprioceptive information about the orientation of the head. This auditory information engages with gain fields in a certain way, or effects a certain set of basis functions, and as a result the subject is in a position, without cognitive preliminary, to walk toward the ball, or orient toward it, or to throw a dart at it -- and thus at the personal level she simply hears that the bowling ball is in front of her in much the same way that you or I can see that the bowling ball is in front of us.[7]

To make the point succinctly and explicitly: the kind of information processing that the PPC engages in (details concerning whether gain fields or basis functions are the right model are not relevant) is to take a sensory signal together with relevant postural signals (eye orientation, head orientation, etc.), and turn this into information usable in action: it is the thing graspable by doing this, the thing towards which one could orient by moving like so. But such behavioral engagements are what the axes and magnitudes of the behavioral space are made of.

I have indulged in this excursion into a theory of the spatial content of perceptual experience in order to lay the groundwork for a theory of the temporal content of experience. This theory will parallel the account of spatial representation in that first I will describe a kind of information-processing framework that supports states that carry temporal information. This information processing framework is an extension of a framework I have been developing and soliciting, the *emulation framework*. The next section, Section 4, provides a very brief introduction to this framework, and Section 5 briefly outlines the extensions to this framework that allow for the production of temporally thick state estimates. Section 6 describes the conditions under which states that carry temporal information, as described in Section 5, have behavioral-temporal *import* for the organism.

4. Representation: The emulation theory

In this section I will provide a *very* quick introduction to a theory of neural representation that I have developed in much more detail elsewhere.[8] The *emulation framework* is an information processing architecture based on constructs from control theory and signal processing. The basic idea is that the brain constructs entities, most likely implemented in systems of neural circuits, that act as models of the body and/or environment. During normal sensorimotor activity the brain sends motor commands to the body, and the body and environment

---

[7] Again, this has been a very brief discussion of a very complicated topic. For much more detail on all of the issues broached in this section, see Grush 1998, 2000, and in preparation.

[8] The best version currently available is Grush 2002. In Grush (in preparation) this framework it yet more fully developed and defended. See also Grush 2003, and Grush 1997 for treatments.

conspire to produce sensory signals sent back to the brain. Circuits which observe the outgoing efferent commands and subsequence incoming afferent signals are positioned to learn the input-output functions of the body and environment. Once such a neural system has learned to some degree of accuracy or precision this mapping, the brain can then use this *emulator* in various ways: it can run the emulator off-line (meaning that the real body and environment are not involved in the control loop) in order to produce imagery; it can run the emulator in parallel with the body and environment and use the output of the emulator to anticipate, fill in, or otherwise enhance the real sensory signal.

The emulation framework begins with neither traditional symbolic computation nor connectionism, but with control theory – there being a clear sense in which the brain is usefully conceived of as, and evolved to be, the controller of the body. The two best-known kinds of control scheme are open-loop and closed-loop (as shown in Figure 2).

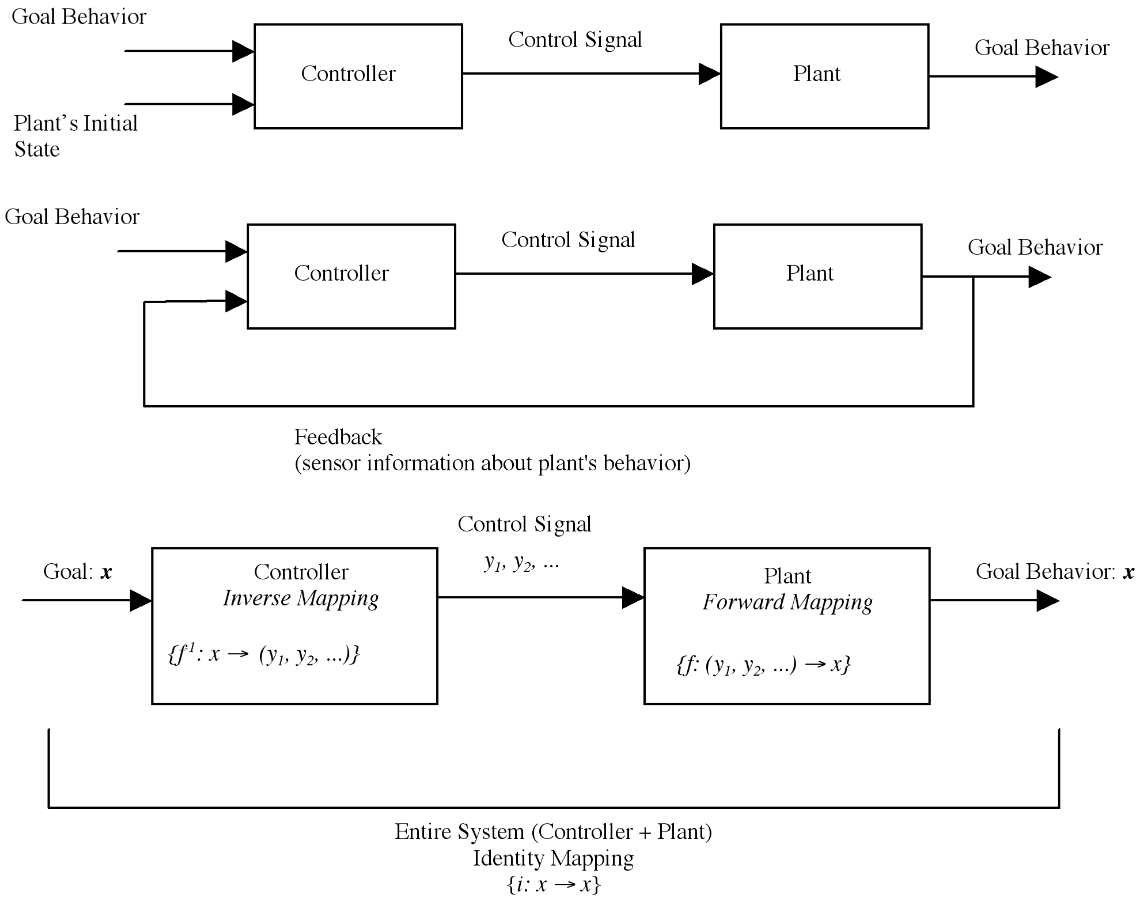

Figure 2: Basic control theory concepts and terminology. The schema at the top is an open-loop scheme, where the controller gets no feedback from the plant. The middle diagram is feedback control, in which the controller exploits feedback from the plant. The bottom diagram explains the notions of inverse and forward mappings. (Figure from Grush, 2002a)

Open-loop schemes take as input a goal state and produce a control sequence such that if the plant executes that sequence, it will end up in the goal state. The controller determines this control sequence in absence of any information about the progress of the plant as it acts on the control sequence. A standard toaster is an example: a goal state (light, medium, dark) is input, and the controller determines how long to keep the heating elements on without any information about the actual state of the bread as it toasts.

A *closed-loop* controller also gets as input a desired goal state and produces a control sequence as output – a control sequence such that if the plant executes it the desired goal will be achieved. Closed-loop controllers, though, have the benefit of feedback from the plant as the control episode progresses. A thermostat is an example of a closed-loop controller: once the desired temperature is set, the controller need not determine the entire control sequence: it need not try to determine at once that it needs to turn the heater on for 44 minutes in order to achieve the desired room temperature. Rather, at each moment it merely compares the current temperature (the feedback in the form of a current temperature reading) with the desired temperature, and takes one of a small number of actions on that basis; e.g. turn heaters on, keep heaters on, turn AC on, etc.

Another, less well-known scheme is pseudo-closed-loop control (aka model reference control, see Figure 3).[9] In such schemes the controlling system includes not only a controller *per se*, but a model or *emulator* of the plant – a system that implements the same input-output mapping as the plant, or close to it. With such a scheme it becomes possible for the controller to operate as if it is in closed-loop contact with the plant even though it is not. A copy of the motor command is sent to the emulator, which, because it implements the same input-output mapping as the plant, produces as output a signal the same as the one the plant is producing. The controller can exploit the emulator's feedback in exactly the same way it would exploit feedback from the real system.

---

[9] The qualitative idea behind this sort of framework was articulated by Kenneth Craik (1943). To my knowledge, it was first put to use in theoretical neurophysiology by Masao Ito (1970, 1984) as a theory of one aspect of cerebellar function.

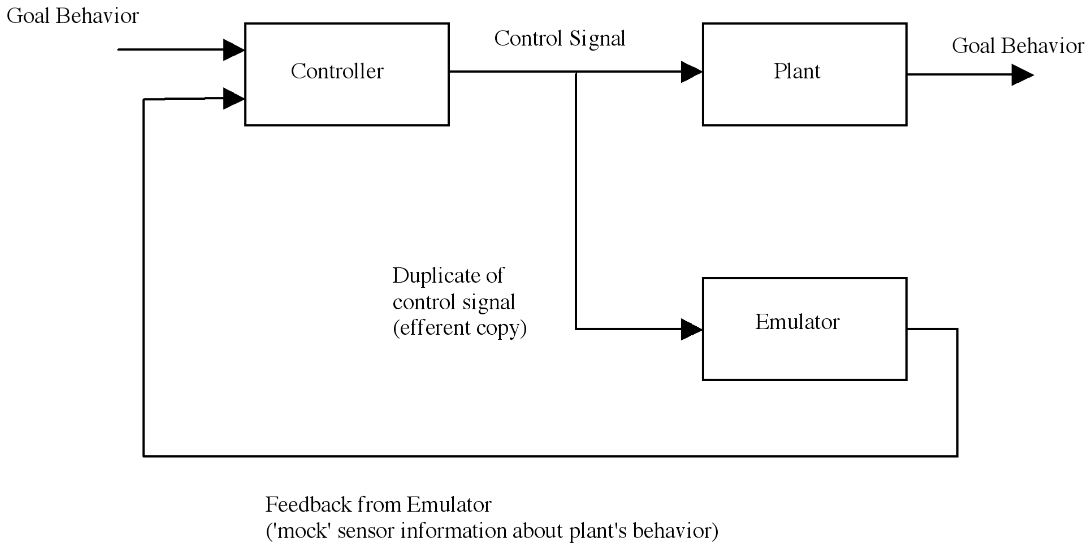

Figure 3. Pseudo-closed-loop control.

There are many potential uses for such a system. I will mention only two. First, the plant can be disengaged completely and the emulator run off-line by the controller in order to produce imagery, and more generally evaluate counterfactuals (if commands sequence C were issued, what would the result be?). Second, in systems where real feedback is compromised or delayed, the emulator can be run in parallel with the plant to provide mock feedback. As a historical note, this was the use proposed by Masao Ito (1970, 1984) as a solution to the problem of delayed proprioceptive feedback: he argued that the cerebellum implements an emulator (not his term) of the musculoskeletal system that provides immediate feedback in order to overcome peripheral feedback delays for quick movements.

In such a scheme an emulator is useful. But in another scheme it can be even more useful. The next step to this more useful scheme is the *Kalman filter*. The Kalman filter (henceforth KF), is an information processing technique for filtering the noise from a certain kind of signal. The KF and the problem it solves are shown in Figure 4.

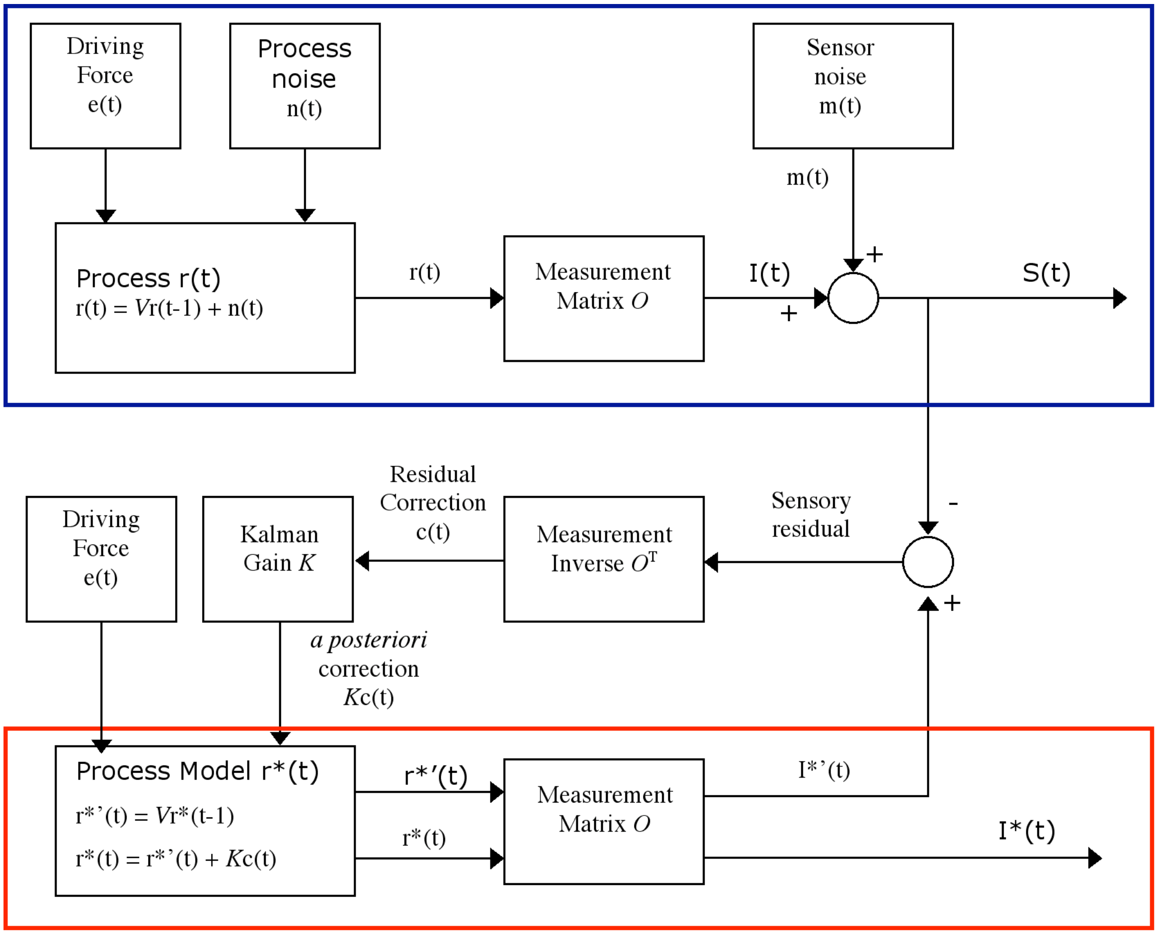

Figure 4. Standard Kalman filter.

The problem is represented on the upper third of Figure 4, outlined in blue. A process r(t) develops over time under three influences. The first is the manner in which the process's states lead to its own future states. This can be expressed as r(t) = *V*r(t-1), where the matrix *V* represents the function that determines future states of *r* on the basis of its past states (this is a Markov process). The second influence is random perturbations, or what is called process noise, represented as a small zero-mean Gaussian 'noise' vector n(t). Thus r(t) = Vr(t-1) + n(t); this would be a Gauss-Markov process. The third influence is an external driving force e(t) that influences the process: r(t) = Vr(t-1) + n(t) + e(t); a *driven* Gauss-Markov process. An example would be a ship at sea. Its states are its position and velocity (speed and direction). At each time step its position, speed and velocity are determined by three things: the position, speed and velocity at the previous time; 'noise' in the form of unpredictable winds or ocean currents, and anything else unpredictable; and driving forces such as the engine speed and rudder angle. Ships at sea can be faithfully described as driven Gauss-Markov processes, as can many other things.

A signal I(t) is produced at each time *t* via a measurement that depends on the state of the process at *t*. To this signal is added some small sensor noise m(t) (again, zero-mean Gaussian), and the result is the *observed signal* S(t). In the ship example the measurement might include bearings to known landmarks and stars (the values produced depend upon the ship's position and orientation). Such bearing measurements are not perfect, and the imperfection can be represented as sensor noise that is added to what 'ideal' perfect bearing measurements would be. Everything just described is represented on the blue-boxed region of Figure 4.

The KF's job is to determine what the real signal I(t) is; that is, to 'filter' the sensor noise m(t) from the observed signal S(t). In the ship example, this would amount to determining what the ship's actual position is, or equivalently what the perfect ideal bearing measurements would be. The KF does this in an ingenious way. It maintains an estimate of the current state of the process, and then measures this *process model* with the same measurement function that produces the real measurement from the real process. Of course, if this state estimate is very accurate, then the KF's estimate of the real signal will be very accurate.

So the trick it to keep the state estimate as accurate as possible. This is done in the following two-phase manner. In the first phase, the KF produces an *a priori* estimate of the real signal, I*'(t), by subjecting it's *a priori* estimate of the state of the process r*'(t) to measurement O. The asterisk indicates that the value is an estimate produced by the KF, and the prime indicates that it is the *a priori* estimate (there will be another, *a posteriori*, estimate produced in the second phase). The KF compares this *a priori* signal estimate to the observed signal. The difference between them is called the *sensory residual*. It is the difference between what the KF *expected* to see and what it *actually* saw.

From this sensory residual, the KF can determine how much it would have to modify its own *a priori* state estimate r*'(t) in order to eliminate the residual. This is the residual correction (and it is determined by pushing the sensory residual through an inverse of the measurement process). Qualitatively, the KF says: I thought the process's state was X, and this led me to predict that the signal would be Y. But the observed signal was actually Y + y. If my state estimate had been X + x, rather than X, then my prediction of the signal would have matched the real signal. Here *y* is the sensory residual, and *x* is the residual correction.

The key point, though, is that the KF does not (typically) alter its state estimate by this amount. Why not? Because the sensory residual -- the mismatch between the *a priori* prediction of what will be observed and what is actually observed -- has two sources. *One* is the difference between the KF's estimate of the process state and the actual state of the process. This is the inaccuracy of the process estimate. The *second* is the sensor noise. That is, the KF's *a priori* state estimate might be entirely accurate, but there would still be a sensory residual because of the sensor noise. And even in the normal case where the *a priori* estimate is not entirely accurate, this inaccuracy is responsible only for part of the sensory residual. To return to the ship example, when the navigation team predicts that they should be at location L, they can also predict what the bearing measurements from the current fix cycle should be if their estimate is accurate. When the actual bearing measurements come in, they will typically not exactly match the navigation team's predictions. One source of the mismatch is that their prediction of where the ship actually is is probably not exactly right. Another source is the imperfection of the bearing-taking process itself.

The KF makes optimal[10] use of the residual correction in the following way. It makes an estimate of its relative confidence in its own estimates versus what the observed signal says, which is largely a matter of the relative size of the process noise and the sensor noise. Upon determining this relative confidence the KF applies, to put it roughly, some fraction of the residual correction to its state estimate. *Very* roughly put, the KF determines a gain matrix, the *Kalman gain*, that determines what fraction of the residual correction to apply.

If the KF knows that the sensor noise is large compared to the process noise (there are few unpredictable winds or ocean currents, for example, but the bearing measurements are known to be of limited reliability), it knows that, probably, most of the sensory residual will be due to inaccurate sensor readings and not to inaccuracies in its own *a priori* state estimate.[11]

---

[10] The use is in fact optimal in the sense that, given the details of the various sources of noise, and the way in which the KF determines the Kalman gain (explained shortly), it is provably the case that it is not possible to get more information from the signal than the KF gets.

[11] This is because process noise is the biggest factor compromising the accuracy of *a priori* estimates. As process noise goes to zero, r*'(t) goes to r(t), so long as the previous state estimate was accurate. As process noise increases, r*'(t) will deviate more and more from r(t), even if the previous state estimate was accurate, because of the unpredictable influence on the process.

It therefore has reason to believe that most of the residual correction should be ignored (it is due to sensor noise). It applies only a small fraction of the residual correction in this case.

On the other hand, if the KF knows that the process noise is large compared to the sensor noise (the instruments are good and those operating them are quite skilled, and furthermore there are large unpredictable winds and ocean currents), then the KF knows that in all likelihood its own *a priori* prediction will not be very accurate, at least not in comparison to what the sensor signal says. So it will apply a large fraction of the residual correction.

Then begins the second phase of the KF's operation. However much of the residual correction gets applied, the result is a new updated estimate of the state of the process that has taken the observed signal into account. This new estimate is the *a posteriori* estimate. This estimate represents an optimal combination of two factors: the (fallible) *a priori* expectation of what the process's state should be, and the (fallible) observed signal. This *a posteriori* estimate is then subjected to a measurement, and the result of this measurement is the KF's final estimate of the real signal -- its estimate of what one would get if one removed the sensor noise from the observed signal.

The process then repeats. The current *a posteriori* estimate is the basis for the next *a priori* estimate.

My introduction to KFs has been extremely incomplete. Suffice it to say that not only is it a provably optimal filter in a broad range of cases, it is one of the most widely used information processing constructs in many areas of signal processing and related areas. The next task will be to integrate the KF into control contexts.

Figure 5 shows a schematic of what, for lack of a better name, I will call the *emulation framework*. It is a combined control and signal-processing framework that blends pseudo-closed-loop control and the Kalman filter. It can be described as a modified closed-loop control scheme that uses a Kalman filter to process the feedback signal from the plant. Or it can be described as a Kalman filter such that the process (together with the measurement process and sources of noise) correspond to the plant, and the driving force is the command signal from the controller. Note that the three control schemes described earlier in the section are all degenerate cases of this scheme: First, if the Kalman gain is set to 1 so that the entire residual correction is always applied, the scheme becomes functionally equivalent to closed loop control. In such a case, the feedback the controller sees will always be identical to the signal sent from the plant, the observed signal. Second, if the Kalman gain is set to null, so that none of the residual correction is applied, the scheme becomes functionally equivalent to the pseudo-closed-loop scheme of Figure 3. In such a case, the feedback sent to the controller will always be exactly what the a priori estimate says it should be, without any influence from the observed signal. Without correction, the state estimate continues to evolve based only on its own estimates, just as in pseudo-closed-loop control. Third, if the feedback signal is suppressed completely, then the scheme becomes functionally equivalent to open-loop control.

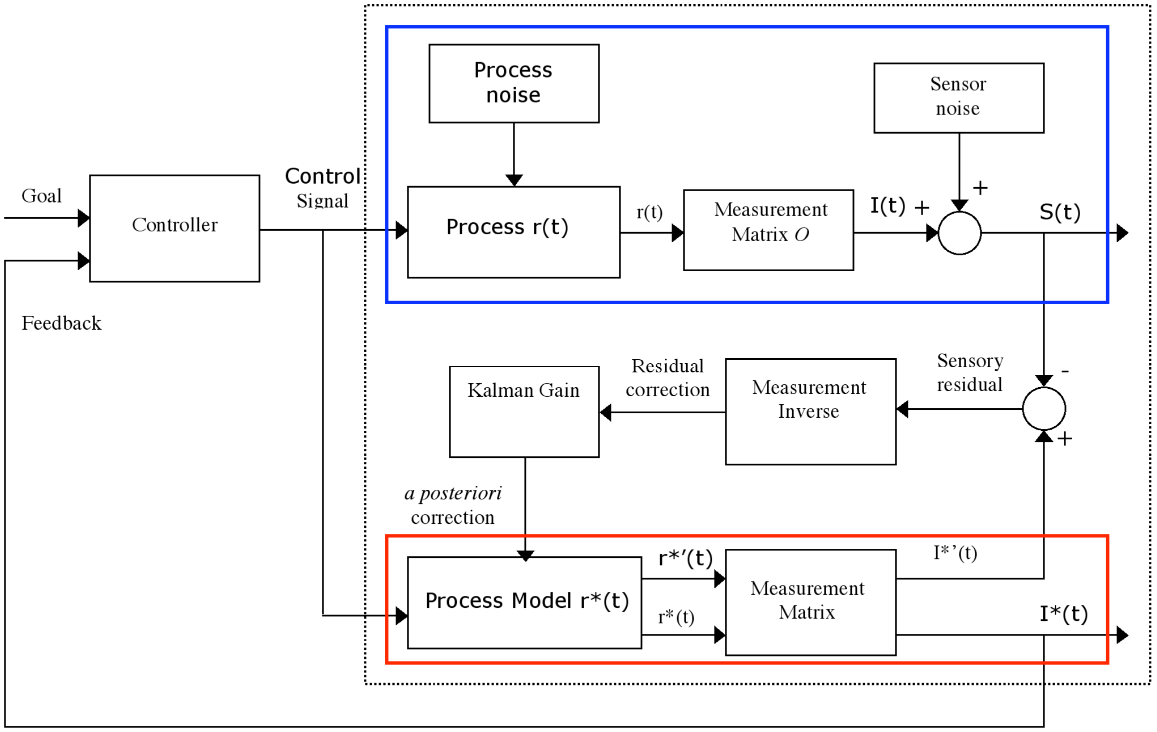

Figure 5. The emulation framework.

The emulation framework is extremely useful for shedding light on phenomena in motor control, motor imagery, visual imagery, the relationship between visual imagery and motor processes, the nature of top-down processes in perception, and a host of others. In motor control, researchers have long known that peripheral proprioceptive feedback is too slow to be of direct use during fast goal directed movements, and some have argued that the brain uses emulators (my expression, not theirs), in KF-control type schemes to maintain an estimate of the state of the body during movement, since the sensory signal is delayed.

I have shown elsewhere how the same mechanisms can explain motor imagery, the imagined feelings of movement and exertion that are the proprioceptive counterparts of visual imagery. This is done, simply enough, by running the process model-emulator but not the body. The emulator thus supplies off-line mock proprioceptive signals.

The same mechanisms can also provide insight into visual imagery. On the current proposal, the brain contains emulators of the motor-visual loop, that provide anticipations of what will be seen on the basis of what is currently being seen and what the current motor commands are. Striking confirmation that such processes are operative comes from a phenomenon first discovered by Helmholtz. When such subjects whose eyes are prevented from moving are shown a scene, and a peripheral stimulus of a sort that would normally cause an eye movement appears (such as a flash), they report seeing the visual scene shift in the direction opposite the stimulus for a brief moment. On the present view, an emulator is given information via an efferent copy that the eyes will be moved, and so an anticipation of the next retinal image is formed to the effect that it will be just like the current retinal image only shifted in the direction opposite the issued eye movement command. During normal overt perception these predictions are largely confirmed and leave no separate conscious trace (this phenomenon will be discussed in more detail below). But since the subjects' eyes don't actually move (having been immobilized), the sensory signal comes back, and the a priori estimate is overruled by the large sensory residual, and the experienced image shifts back.

I have also shown (Grush, 2002), in a way that owes much to Kosslyn, that visual perception employs imagery of this sort, to form anticipations and fill out lacunae. Surely the sort of thing that becomes very salient to the subjects with paralyzed eye muscles happens to us all, but because our eyes actually move, these anticipatory images are largely confirmed and immediately woven into our visual experience.

The applications of this framework could be multiplied, and I have done so elsewhere (see Grush, 2003, in preparation, to appear). For now, the project is to develop an application that I have not discussed anywhere else: temporal representation.

5. Temporal representation: Moving Window Emulation

The Kalman filter, at any given point in time, is actively concerned with only one process state estimate, the current one. It uses the previous state estimate as one of the sources of information when developing the current state estimate, but it does not actively alter or update it once it has been arrived at. And as soon as the previous state estimate is used to construct the current state estimate, it is forgotten – its only trace being its influence on the current state estimate, which will itself be exploited and forgotten when used to construct the subsequent state estimate. The KF has no temporal *depth*. By the same token, the KF is not in the business of predicting *future* states: its current state estimate is use to determine a prediction of what the sensory signal will be, but it is not a prediction of what will be happening in the future with the process. It is a prediction of what will be sensed about the process as it is *right now*. At each t_i, the KF constructs and maintains an estimate of the process' state at t_i. It only had an estimate of t_{i-1} at t_{i-1}, and it will only get around to constructing an estimate for t_{i+1} at t_{i+1}. The KF is therefore temporally punctate in that it represents the states of the process only at one point in time at a time. And because of this it does not need to represent time explicitly at all. It lets time be its own representation (as Carver Mead is widely reported to have said) in that the time of the KF's representing just is the time represented. In this section I will show how to generalize the temporally punctate KF to a system that maintains, at each time, an estimate of the behavior of the process over a temporal interval.

Information processing systems that develop state estimates of past states using information up to the present present are known as *smoothers*.[12] There are different kinds of smoothers depending on the point in the point past that is used to define the interval. *Fixed point smoothers* keep an estimate of the process at some fixed temporal point, and update that estimate as time progresses. Such a system would construct, at each time t_i, ($i>=0$) an estimate of the process's state at t_0. Presumably, as more future information is collected, the estimate of the processes satet at that point in time can be improved.

A *fixed-lag smoother* constructs, at each t_i an estimate of the process' state some fixed amount of time (the lag l) before t_i the past: t_{i-l}. (The proposal by Rao, Eagleman and

---

[12] Of course the same name is applied to things that aren't necessarily temporal in character. By the time this section is over, the analogies between them should be obvious.

Sejnowski is that visual awareness is the result of a fixed-lag smoother, with a lag of about 100-200 msec.)

Information processing systems that produce estimates of *future* states of the process are *predictors*. Where the distance into the future for which a state estimate is produced is *k*, the predictor is a *k-step-ahead predictor*.

With this, we can phrase the filtering problem generally:

Given the entire set of observed signals S_1, S_2, ... S_n, (from the initial time t_1 to time t_n) find, for each t_i, the optimal estimate of the process state r_j.

If i = j, this is filtering (the Kalman filter as described above is thus a filter; below I will distinguish another kind of filter, and will specify the filter as defined here a *moving point filter*);

If j > i, this is prediction, if j = i+k for fixed k, then this is a k-step-ahead predictor. (I will be especially interested in the case where estimates r_i, r_{i+1}, ..., r_{i+k} are maintained for all t_i.)

If j < i, this is smoothing; If j is some fixed point, then this is fixed point smoothing; if j = i - l for some fixed lag l, this is fixed-lag smoothing.

With these definitions in hand, I can define another sort of information processing framework: what I will call a Moving Window Filter (MWF). This is a scheme that combines smoothing, filtering and prediction to maintain, at all t_i, estimates of the process's state for all times within the interval $[t_{i-l}, t_{i+k}]$ for some constant past-lag *l* and some constant future-reach *k*.

One way to express the distinction between a smoother, a bare filter as defined above (what we might now call a *moving-point filter*) and a moving window filter is that at each point in time the moving point filter constructs an optimal estimate of the process *at that time*, while the moving window filter constructs an optimal estimate of the behavior of the process over a temporal interval centered[13] on the current time:

---

[13] It need not be literally centered -- *k* need not equal *l*.

	Fixed Lag (l = 2)	Moving point	Moving Window
t_4:	r_2	r_4	$(r^4_2, r^4_3, r^4_4, r^4_5, r^4_6)$
t_5:	r_3	r_5	$(r^5_3, r^5_4, r^5_5, r^5_6, r^5_7)$
t_6:	r_4	r_6	$(r^6_4, r^6_5, r^6_6, r^6_7, r^6_8)$
t_7:	r_5	r_7	$(r^7_5, r^7_6, r^7_7, r^7_8, r^7_9)$

Let's look first at the column in Table 1 under the moving point filter. The vectors r_4, r_5 etc., are the optimal estimates of the process at t_4, t_5, etc., respectively. But note that there is an ambiguity in the characterization just given. It could mean that it is the estimate, generated at t_4, t_5, of the process's state. Or it could mean that it is the estimate of what the state of the process is at t_4, t_5. With the moving point filter this ambiguity does not make a difference, since the estimate generated at each time step is the estimate of the process's state at that time. Because the time of the estimating is always the same as the time estimated, we can get by with only one index.

Look now at the column for the fixed-lag smoother. Here the ambiguity dissipates -- r_4 is an estimate of the state that the process was in at t_4. But the time that this estimate is generated is t_6. For the fixed-lag smoother, the time represented always lags behind the time of representing by some constant amount. But exactly because this delay is fixed, we don't need separate indices for the time of representing and the time represented. The one index still specifies both.

This is not true for the moving window filter. Because at each time step the filter maintains estimates of the process's state at multiple times, the time of estimating is not always the same as the time estimated, nor is it some fixed temporal distance from the time estimated. Thus the need for two indices on each estimate: one for when the estimate is constructed, another to indicate the time index of the state being estimated. With r^4_2 for example, the superscript index '4' refers to the time of representing, the time at which that estimate is constructed. This is why at t_4 all of the states for the moving window filter have a '4' superscript. The subscript refers to the time represented. So r^4_2 is the estimate generated by the moving window filter at t_4 of what the state of the process was at t_2. The window (r^4_2,

r^4_3, r^4_4, r^4_5, r^4_6) is the filter's estimate, at t_4, of the trajectory of the process from t_2 (in the past) to t_6 (in the future).

One crucial thing to note is that the moving window filter can, and often will, alter its estimates of what the process's state will be/is/was as time progresses. That is, r^4_4 need not equal r^5_4 even though they are both estimates of what the process's state is at t_4. The MWF's estimate at t_4 of the process's state *is* at t_4 might not be the same as its estimate at t_5 of the process's state *was* at t_4. The MWF might decide, on the basis of incoming information at t_5 that it had made an error. That while at t_4 it had thought that the state of the process was r^4_4, it now, at t_5, estimates that this was probably not correct. And so it now, at t_5, retroactively alters its estimate of what the process's state was at t_4.

For a qualitative example, imagine that we have a ship navigation team that employs the moving window filter technique. At each time the team maintains an estimate of the ship's trajectory from two fix-cycles ago to two-fix cycles in the future. The estimates of the previous and current ship states (position, speed, bearing, etc.) are made on the basis of sensor information that has been received. There are two kinds of situation that would lead the team to retroactively modify its state estimates. First, suppose that at t_3 the best information available had the ship's state as being r^3_3. However, t_4 at the new information comes in that strongly suggest that the ship's current state is r^4_4. But suppose that because of the nature of the process, if the ship's state at t_4 is , then the ship could not have been in r_3 at t_3. The team can thus use this information concerning the ship's state at t_4 to correct its estimate of where it *was* at t_3. In such a case r^3_3 (the estimate, at t_3, of the ship's state at t_3) would differ from r^4_3 (the estimate, at t_4, of the ship's state at t_3).

A second kind of situation is worth mentioning. Suppose that the ship's measurement systems -- the people who take bearings and report the numbers to the navigation team -- are spread throughout the ship, and their measurements are carried by hand to the navigation team. While all or most of the information gets to the navigation team, some of it is significantly delayed because the ship is large and the couriers are slow. At t_3 the navigation team estimates that the ship's current state is r^3_3. At the next fix cycle, t_4, a courier comes in with some bearing measurements that were taken at t_3. This information can then be used to modify the team's estimate of the ship's state at t_3. Again, $r^3{}_3$ will differ from $r^4{}_3$, but this time because of the incorporation of delayed information, not because of consistency constraints between present and past state estimates.

6. Temporal import

Information processing systems of the sort just described provide examples of how a physical systems can have states that, at a given time, can carry information about states of other systems at times other than the present time. But as we saw in the case of spatial representation, a person's brain having states that carry X-ish information is not sufficient for that person having experience with X-ish import. In the case of space, recall, the proposal was that it was a matter of the information-bearing states being appropriately mainlined into the organism's sensorimotor repertoire that imbued those states with spatial content. To recap: the kind of spatial content that we wanted to explain was in terms of the behavioral space, and the axes and magnitudes of the behavioral space are defined in terms of the organism's behavioral capacities. The proposal was that the PPC, exactly because it is in the business of putting sensory information (together with postural information) into a format that can cue and guide motor behavior, imbues those sensory episodes with behavioral spatial import.

The proposal for the case of temporal import will be parallel. What makes a given information-carrying state support experience with content specifiable in behavioral time will be the fact that that state plays a role in the right sensorimotor dispositions, in this case, the temporal features of such dispositions.

But it will be convenient to pitch the proposal using as an example something simpler than the information-carrying states supplied by a moving-window filter. Suppose that an experimental condition is set up such that a small red point of light is quickly flashed on a subject's retina exactly 700ms before an irritating itch, that will last a few seconds, occurs on her forearm. The flash thus carries temporally conditioned information, but its initial import will simply be of a quick flash of red light, lacking any interesting temporal import (except import to the effect that a red light is flashing *now*). Now I suppose that after a while the subject might come to associate the red flash with the impending itch, and perhaps such an association would be enough to give the flash some kind of temporal significance.

Rather than pursue that thin hope, though, I want to upgrade the example so that the light has specific behavioral ramifications. The experimental set-up is such that if the subject scratches a specific spot on her forearm not earlier than 600ms after the red light onset and not later than 650ms after onset, then the itch will be prevented. Given these constraints the timing will be quite tricky, and we can suppose that when informed of the conditions the subject will initially take the flash of light as a cue to time her motor response. This would be parallel to you or I donning the sonic guide and, being informed that Middle C at 35 dB carries information to the effect that a bowling ball is on the floor two meters ahead, and then trying to hit it with a dart by aiming at a spot we take to be about 2 meters ahead.

But as we saw in the case of the sonic guide, an appropriately skilled subject has a different phenomenology than this. To her, that same auditory input results in a perceptual state that is best described as ‘perceiving the ball just ahead of her’ or something like that. Her phenomenology is different from ours, though the sensory inputs are the same. My *claim* is that our subject will, after sufficient experience, undergo a similar phenomenological change. The red light will stop having the import of ‘red light in my visual field now right there’ (where the ‘there’ refers to a spot in her visual field), but will come to have the import ‘itch *there* just *then’*. There ‘there’ will have as its import the location on her forearm, and the ‘then’ will have as its import a specific temporal location in the subject’s behavioral time, an anticipated point in the immediate future. A point in time exactly as phenomenally real as the point in the future at which you anticipate your hand catching the pencil as it rolls off the edge of the desk.

That anyway is the claim. Aside from the sonic guide analogy I have not defended it, and I won’t defend it here since that would take some time, and I’m not sure I could do a good job anyway. But though it is a speculation, I should point out that to my knowledge it is the only speculation going. The fact is that because phenomenological time is not the same as real, objective time, there has to be some account of temporal phenomenology. A phenomenology that flows through time is not the same as temporal phenomenology (experimental results confirm this, some of which have been mentioned, and others will be mentioned briefly), and so the notion that we don't need to account for temporal phenomenology, but can instead just rely on the assumption that time is its own representation, is untenable. While the speculation here might not ultimately be correct, the problem it is an attempt to address won't go away by ignoring it or failing to recognize it.

7. Discussion

7.1 Comparison with two recent proposals

There are two psychological phenomena that bring out clearly the need for some sort of theory of temporal representation. The first is Geldard and Sherrick's cutaneuous rabbit. These researchers placed tactile stimulators that deliver a small thump on subject's arms: one near the wrist, one near the elbow, and one near the shoulder. A series of taps is delivered to the subject with very brief interstimulus intervals, on the order of 50-200ms. A series of 5 or 10 taps might take anywhere from a fraction of a section to a couple of seconds. If, say, four taps are delivered to the wrist, and nothing else, the subject will feel and report a series of taps all located on the wrist. But if the series delivered is, say four taps on the wrist, three at the elbow and three at the shoulder, the subject will feel, and report, a series of *evenly spaced taps*, like the footsteps of a small animal running up the arm from the wrist to the shoulder. The puzzle begins with the second tap on the wrist. If in the future there will be some taps, say taps five through seven, located near the elbow, and maybe a few more after that near the shoulder, then the second tap will be felt a few inches up from the wrist in the direction of the elbow. If on the other hand, the taps will not continue beyond the 4th tap at the wrist, then the second tap will be felt at the wrist. But how does the brain know, at the time the second tap is felt, whether or not taps will be delivered near the elbow later on? Obviously, it doesn't.

The second is the finding concerns the flash-lag effect (MacKay, 1958). The phenomenon is that subjects are shown a scene in which a moving object and small flash occur at the same place at the same time, but the moving object appears to be ahead of the flash. (illustration, moving ring and flash circle). The finding of interest is that the magnitude and even direction of the flash-lag effect can alter depending on what the moving object does *after* the flash (see Eagleman and Sejnowski 2000; Rao, Eagleman and Sejnowski 2001).

With phenomena like these in mind, Daniel Dennett (Dennett 1992; Dennett and Kinsbourne 1992) has proposed what he calls the 'multiple drafts' model. According to this model, the brain is not a passive mirror of sensory input, but is rather an active interpreter, and in particular in interpreter that is not bashful about re-interpreting (hence the 'multiple drafts' metaphor) its previous interpretations of perceptual states based on subsequent information. I will return to Dennett and Kinsbourne shortly

The model proposed by Rao, Eagleman and Sejnowski in order to account for these results is that the visual system is a fixed-lag smoother, as described in Section 5. As they put it:

> In the smoothing model, perception of an event is not online but rather is delayed, so that the visual system can take into account information from the immediate future before committing to an interpretation of the event.

As eagleman has elsewhere put it, we are living in the past, by around 100 msec or so.

Though Eagleman and Sejnowski cite Dennett as a theoretical ally, their model is closer to the Cartesian Theatre that Dennett criticizes than to a Dennettian multiple drafts model. On their model, there is one definitive percept -- one draft, so to speak, not multiple drafts. The departure from the Cartesian Theatre model as Dennett presents it is that this single definitive percept is delayed by a hundred milliseconds or so, and what is shown on the Cartesian screen as having occurred at *t* depends in part on the nature of the sensory input that comes in after that sensory input that is most directly implicated in carrying information about what happened at *t*. But a theater performance that starts 100 msec late in order to accommodate a tardy script writer is still a theatre performance, even if the reason that the script writer is tardy is that he wants his script to be consistent with things that take place after the time of the performance.

The moving-window emulation theory is similar in spirit to Rao, Eagleman and Sejnowski's fixed-lag smoother model. It has all advantages and lacks the shortcomings. First, to the advantages it shares. Rao et al.'s fixed-lag smoother applies its interpretive energies at the endpoint of the relevant temporal window, the fixed lag. This allows such a system to exploit, in its construction of percepts, information that has arrived after the event the percept concerns. This is why they proposed the model, after all. Obviously the moving window emulator does the same. The trailing edge of the window is a fixed lag, and at this trailing edge and beyond, the two models are functionally equivalent. For an event that occurs at *t*, both models predict that from *t* + *lag* onward, the percept associated with that event is influenced by what happened after *t*, up to and including *t* + *lag*.

Now to the disadvantages that the moving window emulator does not have. The problem with fixed-lag smoothers is that they delay their estimates, and in sensorimotor contexts, this can be costly. A fixed-lag smoother must balance these costs against the gains made by having a more accurate (though delayed) percept. The moving window emulator does not need to delay its processing. In fact, it even engages in anticipatory estimates before the event. While this has the theoretical advantage that it avoids the disadvantages of delaying the percept, it also seems to accord with empirical results that appear to indicate that anticipatory percepts are in fact constructed (e.g. the Helmholtz phenomenon, but more generally visual perception appears to exploit a wide range of anticipatory mechanisms).

The second avoided disadvantage is connected with the first. On the fixed-lag model, our phenomenology is necessarily epi-phenomenal. Our brain and its mechanisms do the needed sensorimotor work, and then after the fact our phenomenology shows up, and provides us with the delayed illusion that we were in control. While the mere fact that this is counter-intuitive is no decisive argument against it -- cognitive neuroscience has demonstrated that many counter-intuitive theses are true -- it should at least motivate us to take seriously any contenders that that all the advantages but lack the counter-intuitive consequences. The moving window emulation theory does not entail that our phenomenal experience is epiphenomenal, not does it entail that it is not. Because on the model there are state estimates (percepts) concerning what is happening at *t* available at *t*,(and maybe in some cases even a bit before t, as with anticipatory percepts) it remains possible that our percepts are genuine players in our sensorimotor engagements.

Finally, (and this may or may not ultimately be an advantage) the model here is much more consistent with Dennett and Kinsbourne's multiple drafts model. In the first place, there really are *multiple* drafts. The moving window emulator constructs estimates for the entire trajectory at each time, and is in the business of re-writing many of these estimates as a function of anticipations, surprising input, or consistency constraints.

The primary advantage of the moving window emulation model over the multiple drafts model is clarity. It allows us to cash in the loose copy-editing metaphor for some mathematically clear apparatus from control theory and signal processing.

Though this section has been to some extent critical of Dennett and Kinsbourne and also Rao, Eagleman and Sejnowski, I hope it is recognized that this is friendly criticism. By the lights of the model I have articulated, these researchers are, more than anyone else, on the right track. It is because their views are so close to my own that it is useful to clarify my own proposal by highlighting the points of divergence. I am hopeful that these researchers will see my proposal not as a adversarial competitor, but as a friendly successor. It captures everything that Dennett and Kinsbourne wanted to capture with their Multiple Drafts model, at least for phenomena within the window of the behavioral now, but replaces the metaphor with clear mathematical apparatus. And it captures everything that Rao et al. wanted: a filtering mechanism that accommodates the phenomenon of perceptual post-diction. But the model here is simply more general in that it is not limited to post-diction, but is capable of addressing a wider range of phenomena.

7.2 How big is the window?

The information processing structure just introduced shows how a system can maintain, at a given time, an estimate of what is occurring with some target system over an interval. But if this is in fact what the brain is doing, the questions arise: Why is there a window? How big is the window? What is the lag of the trailing edge of the window, and what is the reach of the leading edge? Eagleman and Sejnowski's data suggests that the trailing edge is about 100 msec or so. They have an explanation for why there is this lag, and why it is the magnitude it is. The reason, they suggest, is that it is beneficial to have more accurate percepts, and because the smoother uses information that comes in after the event to interpret the event, the percept will typically be more accurate. The reason that the lag has the magnitude it has is that this particular length of lag represents an optimal (or at least very good) trade off between the benefit of having more accurate estimates and having those estimates delayed. As I have already pointed out, this reasoning highlights a theoretical advantage of the moving window filter over the fixed-lag smoother -- the fixed lag smoother must delay coming to a decision about what happened, and this has clear disutility, as Eagleman et al.'s analysis recognizes.

But the proponent of the moving window filter cannot make use of this form of reasoning, because the moving window filter does not have this cost. So why, according to the moving window account, is there a window, and why does it have the magnitude that it has?

There are facts that provide a possible explanation for why there should not only be a lag, but a reach, of some particular length: feedback delays. First to lag. At a given time *t*, the sensory receptors of an organism are in some state of stimulation that comprise the sensory snapshot of the organism's state. But due to the fact that these signals are carried by channels with finite information transmission velocity, some parts of that snapshot will not be centrally available until after some time has passed. Call the typically maximum sensory delay *d*. In the case of humans, this is probably proprioceptive information from the extremities, and its maximum delay is probably on the order of one to two hundred milliseconds at most. If the central nervous system wants to have an accurate central representation of what is happening or has happened at any time, then it needs to have its representation of what is happening at a given time *t* remain open to revision for a time at least as long as *d*, for the obvious reason that information relevant to that state estimate might be delayed as long as *d*. Clearly, then, this is a theoretically motivated reason for an organism's CNS to adopt *d* as the minimum lag of the trailing edge of its MWF. But since *d* is the typical *maximum* delay of the sensory information's arrival at the CNS, there won't be any strong reason to make the lag longer than *d*, especially since increasing the lag is likely to be of non-trivial cost in terms of neural computational resources.

What about the reach of the leading edge of the MWF? Again, perhaps surprisingly, feedback delay provides a possible answer. According to the moving window emulation account, one of the things the CNS does is to continually generate predictions of what is going to happen. How far into the future should the CNS generate predictions? The major class of predictive phenomena are those based on the CNS's own motor commands (this is what the Helmholtz phenomenon is: the current eye movement motor command generates an expectation as to what the retinal image will change over the next few hundred milliseconds). But motor commands are subject to a delay in execution just as sensory states are subject to a delay. Call the maximum typical motor command delay d_m. Given this, the CNS at *t* has the information it needs to generate a predictions of what will be happening (as a function of its own motor commands) up to $t + d_m$. This is, so to speak, how far into the future the CNS's causal influence reaches, and thus it sets a minimum desirable reach for the leading edge of the prediction side of the MWF.

While these considerations are obviously not watertight, they do provide some *prima facie* motivation for the CNS to have evolved to exploit a MWE information-processing structure, and why the window involved might have some specific magnitude.

7.3 Towards a Kantian theoretical cognitive neuroscience

Those who are involved in cognitive neuroscience and who have an opinion on Kant are typically not fans. And Kant certainly thought that the project of trying to understand the mind by studying the brain was ill-conceived and confused.[14] This latter fact is no doubt part of the explanation of the former. Such postures notwithstanding, cognitive neuroscience stands to gain much by exploring various Kantian lines of thought.[15] While both Hume and Kant conceived of the mind as quite active in its processing of sensory information, Hume conceived of this activity as essentially reconfiguring and recombining, in perhaps novel ways, materials that have been gained exclusively through perception. Kant, on the other hand, argued that at least some of the material with which the mind worked were not provided to it through perception, but were supplied by the mind itself to perception.

The basic idea is represented quite clearly in the two quotes with which I headed this article. Hume takes it that the mind's ideas of time and duration are no more than the temporal features of what is provided to it in sensation: our impression of succession comes from a succession of impressions. For Kant, though, the time (and space) that we experience are constructed by the mind, not given in experience. Space and time, as forms of intuition, are the structures within which the mind interprets its experience, and hence time as experienced is not read off of mind-independent objects or sensation, but is added by the mind in its interpretation of what it perceives.

The theoretical moves by Dennett and Kinsbourne, and Eagleman and Sejnowski, are nods in Kant's direction here. Since the real world proceeds one instantaneous moment at a time, and once a time has past, it has past, the only way that we could experience things at times or in orders other than those in which they objectively occur is if our experience of time is due not

---

[14] See, e.g., Kant's preface to neuroanatomist Thomas Sommerring's *Über das Organ der Seele* (1795).

[15] I am not the first to push this point. Patricia Kitcher's *Kant's Transcendental Psychology*, and Andrew Brook's *Kant and the Mind* are recent examples from the philosophical end. I should mention that though I share the large-scale view of these authors, their interpretation of Kant and their views on how to best understand a Kantian contribution to cognitive science differ from my own on many specifics.

to the temporal features of the things themselves, but due to the brain's own interpretation. And time is not the only such element: for Kant, the mind-supplied contents included space, objects, and the perception of causal relations. And as in the case of time, various results and theories in cognitive neuroscience are beginning to suggest that these elements are, as Kant maintained, interpretive elaborations supplied by the mind/brain, and not contents merely received from without. The present theory is best viewed as an attempt to contribute to this larger project of helping to move cognitive neuroscience from its Humean phase to a Kantian one.[16]

The challenge, of course, is to explain where these mind-supplied elements come from if not the world itself, and how they have the content that they do. It's not magic, and so the accounting has to be done somehow. This is where cognitive neuroscience should part company with Kant. We cannot let these puzzles force us to recognize another realm beyond the one science studies, either as Kant himself did, or as some of the scientists who have worked on temporal experience seem to be willing to allow (e.g. Libet). This is not an easy undertaking.

Acknowledgements:

I would like to thank audiences at the Carleton Philosophy and Neuroscience Conference and UCSD's Experimental Philosophy Lab, where a progenitor of this paper was presented in October of 2002, for helpful discussion and feedback.

References (very incomplete)**:**

Dennett, Daniel (1992). *Consciousness Explained*.

Dennett, Daniel, and Kinsbourne, Marcel (1992). Time and the observer. *Behavioral and Brain Sciences*. 15(2):183-247.

Eagleman and Sejnowski (@). Motion integration and postdiction in visual awareness. Science 287:2036-2038.

Geldard, F.A, and Sherrick, C.E. (1972). The Cutaneous 'rabbit': a perceptual illusion. Science 178:178-9.

---

[16] I am here echoing Wilfrid Sellars' suggestion that his own work was an attempt to move analytic philosophy from its Humean phase to a Kantian one.

Grush, Rick (2000). Self, world and space: on the meaning and mechanisms of ego- and allo-centric spatial representation. Brain and Mind 1(1):@@

Grush, Rick (to appear). The emulation theory of representation: perception, imagery and motor control. *Behavioral and Brain Sciences*.

Grush, Rick (in preparation). *The Machinery of Mindedness*.

Hume, David (@). A Treatise of Human Nature.

Husserl, Edmund (1905). *The Phenomenology of Internal Time-Consciousness*.

Kant, Immanuel (@). Prolegomena to any future metaphysics.

Rao, Eagleman and Sejnowski (@). Optimal smoothing in visual motion perception. Neural Computation 13:1243-1253.

Stern, W. (1897). Psychische Präsenzeit. *Zeitschrift für Psychologie,* 13:325ff.