BEHAVIORAL AND BRAIN SCIENCES (1992) 15, 183–247
Printed in the United States of America

Time and the observer: The where and when of consciousness in the brain

Daniel C. Dennett[a] and Marcel Kinsbourne[b]
[a]*Center for Cognitive Studies, Tufts University, Medford, MA 02155;*
[b]*Behavioral Neurology Unit, Sargent College, Boston University, Boston, MA 02215*
Electronic mail: *ddennett@pearl.tufts.edu*

Abstract: We compare the way two models of consciousness treat subjective timing. According to the standard "Cartesian Theater" model, there is a place in the brain where "it all comes together," and the discriminations in all modalities are somehow put into registration and "presented" for subjective judgment. The timing of the events in this theater determines subjective order. According to the alternative "Multiple Drafts" model, discriminations are distributed in both space and time in the brain. These events do have temporal properties, but those properties do not determine subjective order because there is no single, definitive "stream of consciousness," only a parallel stream of conflicting and continuously revised contents. Four puzzling phenomena that resist explanation by the Cartesian model are analyzed: (1) a gradual apparent motion phenomenon involving abrupt color change (Kolers & von Grünau 1976), (2) an illusion of an evenly spaced series of "hops" produced by two or more widely spaced series of taps delivered to the skin (Geldard & Sherrick's "cutaneous rabbit" [1972]), (3) backwards referral in time, and (4) subjective delay of consciousness of intention (both reported in this journal by Libet 1985a; 1987; 1989a). The unexamined assumptions that have always made the Cartesian Theater so attractive are exposed and dismantled. The Multiple Drafts model provides a better account of the puzzling phenomena, avoiding the scientific and metaphysical extravagances of the Cartesian Theater: The temporal order of subjective events is a product of the brain's interpretational processes, not a direct reflection of events making up those processes.

Keywords: consciousness; discrimination; illusion; localization; memory; mental timing; perception; subjective experience

> I'm really not sure if others fail to perceive me or if, one fraction of a second after my face interferes with their horizon, a millionth of a second after they have cast their gaze on me, they already begin to wash me from their memory: forgotten before arriving at the scant, sad archangel of a remembrance.
> Ariel Dorfman, *Mascara*, 1988

When scientific advances contradict "common sense" intuitions, the familiar ideas often linger on, not just outliving their usefulness but even confusing the scientists whose discoveries ought to have overthrown them. Diagnosed here is a ubiquitous error of thinking that arises from just such a misplaced allegiance to familiar images, illustrated with examples drawn from recent work in psychology and neuroscience. Although this is a "theoretical" paper, it is addressed especially to those who think, mistakenly, that they have no theories and no need for theories. We show how uncontroversial facts about the spatial and temporal properties of information-bearing events in the brain require us to abandon a family of entrenched intuitions about "the stream of consciousness" and its relation to events occurring in the brain.

In Section 1, we introduce two models of consciousness, the standard Cartesian Theater and our alternative, the Multiple Drafts model, briefly describing four phenomena of temporal interpretation that raise problems for the standard model. Two of these, drawn from the research of Libet, have been extensively debated on methodological grounds, but concealed in the controversy surrounding them are the mistaken assumptions we expose. In Section 2, we diagnose these intuitive but erroneous ideas and exhibit their power to create confusion in relatively simple contexts. We demonstrate the superiority of the Multiple Drafts model of consciousness by showing how it avoids the insoluble problems faced by versions of the Cartesian Theater. In Section 3, we show how covert allegiance to the Cartesian Theater has misled interpreters of Libet's phenomena and how the Multiple Drafts model avoids these confusions.

1. Two models of consciousness

1.1. Cartesian materialism: Is there a "central observer" in the brain? Wherever there is a conscious mind, there is a *point of view*. A conscious mind is an observer who takes in the information that is available at a particular (roughly) continuous sequence of times and places in the universe. A mind is thus a *locus of subjectivity*, a thing it is like something to be (Farrell 1950; Nagel 1974). What it is like to be that thing is partly determined by what is available to be observed or experienced along the trajectory through space-time of that moving point of view, which for most practical purposes is just that: *a point*. For instance, the startling dissociation of the sound and appearance of distant fireworks is explained by the different

 0140-525X/92 $5.00+.00

transmission speeds of sound and light, arriving *at the observer* (at that point) at different times, even though they left the source simultaneously. But if we ask where precisely in the brain that point of view is located, the simple assumptions that work so well on larger scales of space and time break down. It is now quite clear that there is no single point in the brain where all information funnels in, and this fact has some far from obvious consequences.

Light travels much faster than sound, as the fireworks example reminds us, but it takes longer for the brain to process visual stimuli than to process auditory stimuli. As Pöppel (1985/1988) has pointed out, thanks to these counterbalancing differences, the "horizon of simultaneity" is *about* 10 meters: Light and sound that leave the same point about 10 meters from the observer's sense organs produce neural responses that are "centrally available" at the same time. Can we make this figure more precise? There is a problem. The problem is not just measuring the distances from the external event to the sense organs, or the transmission speeds in the various media, or allowing for individual differences. The more fundamental problem is deciding what to count as the "finish line" in the brain. Pöppel obtained his result by comparing behavioral measures: mean reaction times (button-pushing) to auditory and visual stimuli. The difference ranges between 30 and 40 msec, the time it takes sound to travel approximately 10 meters (the time it takes light to travel 10 meters is only infinitesimally different from zero). Pöppel used a peripheral finish line – external behavior – but our natural intuition is that the *experience* of the light and sound happens *between* the time the vibrations strike our sense organs and the time we manage to push the button to signal that experience. And it happens somewhere *centrally*, somewhere in the brain on the excited paths between the sense organ and muscles that move the finger. It seems that if we could say exactly *where* the experience happened, we could infer exactly *when* it happened. And vice versa: If we could say exactly when it happened, we could infer where in the brain conscious experience was located.

This picture of how conscious experience must sit in the brain is a natural extrapolation of the familiar and undeniable fact that *for macroscopic time intervals*, we can indeed order events into the categories "not yet observed" and "already observed" by locating the observer and plotting the motions of the vehicles of information relative to that point. But when we aspire to extend this method to explain phenomena involving very short intervals, we encounter a *logical* difficulty: If the "point" of view of the observer is spread over a rather large volume in the observer's brain, the observer's own subjective sense of sequence and simultaneity *must* be determined by something other than a unique "order of arrival" because order of arrival is incompletely defined until we specify the relevant destination. If A beats B to one finish line but B beats A to another, which result fixes subjective sequence in consciousness (cf. Minsky 1985, p. 61)? Which point or points of "central availability" would "count" as a determiner of *experienced* order, and why?

Consider the time course of normal visual information processing. Visual stimuli evoke trains of events in the cortex that gradually yield content of greater and greater specificity. At different times and different places, various "decisions" or "judgments" are made: More literally, parts of the brain are caused to go into states that differentially respond to different features, for example, first mere onset of stimulus, then shape, later color (in a different pathway), motion, and eventually object recognition. It is tempting to suppose that there must be some place in the brain where "it all comes together" in a multimodal representation or display that is *definitive* of the content of conscious experience in at least this sense: The temporal properties of the events that occur in that particular locus of representation determine the temporal properties – of sequence, simultaneity, and real-time onset, for instance – of the subjective "stream of consciousness." This is the error of thinking we intend to expose. Where does it all "come together?" The answer, we propose, is nowhere. Some of the contentful states distributed around in the brain soon die out, leaving no traces. Others do leave traces, on subsequent verbal reports of experience and memory, on "semantic readiness" and other varieties of perceptual set, on emotional state, behavioral proclivities, and so forth. Some of these effects – for instance, influences on subsequent verbal reports – are at least symptomatic of consciousness. But there is no one place in the brain through which all these causal trains must pass to deposit their contents "in consciousness" (see also Damasio 1989a).

The brain must be able to "bind" or "correlate" and "compare" various separately discriminated contents, but the processes that accomplish these unifications are themselves distributed, not gathered at some central decision point, and as a result, the "point of view of the observer" is spatially smeared. If brains computed at near the speed of light, as computers do, this spatial smear would be negligible. But given the relatively slow transmission and computation speeds of neurons, the spatial distribution of processes creates significant temporal smear – ranging, as we shall see, up to several hundred milliseconds – within which range the normal commonsense assumptions about timing and arrival at the observer need to be replaced. For many tasks, the human capacity to make conscious discriminations of temporal order drops to chance when the difference in onset is on the order of 50 msec (depending on stimulus conditions), but this variable threshold is the result of complex interactions, not a basic limit on the brain's capacity to make the specialized order judgments required in the interpretation and coordination of perceptual and motor phenomena. We need other principles to explain the ways *subjective temporal order* is composed, especially in cases in which the brain must cope with rapid sequences occurring at the limits of its powers of temporal resolution. As usual, the performance of the brain when put under strain provides valuable clues about its general modes of operation.

Descartes, early (1664) to think seriously about what must happen inside the body of the observer, elaborated an idea that is superficially so natural and appealing that it has permeated our thinking about consciousness ever since and permitted us to defer considering the perplexities – until now. Descartes decided that the brain *did* have a center: the pineal gland, which served as the gateway to the conscious mind. This was the only organ in the brain that was in the midline, rather than paired, with left and right versions. The pineal looked different, and because its function was then quite inscrutable (and still

is), Descartes posited a role for it: For a person to be conscious of something, traffic from the senses had to arrive at this station, where it thereupon caused a special – indeed magical – transaction to occur between the person's material brain and immaterial mind. When the conscious mind then decided on a course of bodily action, it sent a message back "down" to the body via the pineal gland. The pineal gland, then, is like a theater in which information is displayed for perusal by the mind.

Descartes' vision of the pineal's role as the turnstile of consciousness (we might call it the Cartesian bottleneck) is hopelessly wrong. The problems that face Descartes' interactionistic dualism, with its systematically inexplicable traffic between the realm of the material and the postulated realm of the immaterial, were already well appreciated in Descartes' own day, and centuries of reconsideration have only hardened the verdict: The idea of the Ghost in the Machine, as Ryle (1949) aptly pilloried it, is a nonsolution to the problems of mind. But whereas materialism of one sort or another is now a received opinion approaching unanimity,[1] even the most sophisticated materialists today often forget that once Descartes' ghostly *res cogitans* is discarded, there is no longer a role for a centralized gateway, or indeed for any *functional* center to the brain. The brain itself is Headquarters, the place where the ultimate observer is, but it is a mistake to believe that the brain has any deeper headquarters, any inner sanctum, arrival at which is the necessary or sufficient condition for conscious experience.

Let us call the idea of such a centered locus in the brain *Cartesian materialism*, because it is the view one arrives at when one discards Descartes' dualism but fails to discard the associated imagery of a central (but material) theater where "it all comes together." Once made explicit, it is obvious that this is a bad idea, not only because, as a matter of empirical fact, nothing in the functional neuroanatomy of the brain suggests such a general meeting place, but also because positing such a center would apparently be the first step in an infinite regress of too-powerful homunculi. If all the tasks Descartes assigned to the immaterial mind have to be taken over by a "conscious" *sub*system, its own activity will either be systematically mysterious or decomposed into the activity of further subsystems that begin to duplicate the tasks of the "nonconscious" parts of the whole brain. Whether or not anyone explicitly endorses Cartesian materialism, some ubiquitous assumptions of current theorizing presuppose this dubious view. We show that the persuasive imagery of the Cartesian Theater, in its materialistic form, keeps reasserting itself, in diverse guises, and for a variety of ostensibly compelling reasons. Thinking in its terms is not an innocuous shortcut; it is a bad habit. One of its most seductive implications is the assumption that a distinction can *always* be drawn between "not yet observed" and "already observed." But, as we have just argued, this distinction *cannot* be drawn once we descend to the scale that places us within the boundaries of the spatiotemporal volume in which the various discriminations are accomplished. Inside this expanded "point of view," spatial and temporal distinctions lose the meanings they have in broader contexts.

The crucial features of the Cartesian Theater model can best be seen by contrasting it with the alternative we propose, the Multiple Drafts model:

All perceptual operations, and indeed all operations of thought and action, are accomplished by multitrack processes of interpretation and elaboration that occur over hundreds of milliseconds, during which time various additions, incorporations, emendations, and overwritings of content can occur, in various orders. Feature-detections or discriminations *have to be made only once.* That is, once a localized, specialized "observation" has been made, the information content thus fixed does not have to be sent somewhere else to be *re*discriminated by some "master" discriminator. In other words, it does not lead to a re-*presentation* of the already discriminated feature for the benefit of the audience in the Cartesian Theater. How a localized discrimination contributes to, and what effect it has on the prevailing brain state (and thus awareness) can change from moment to moment, depending on what else is going on in the brain. Drafts of experience can be revised at a great rate, and no one is more correct than another. Each reflects the situation at the time it is generated. These spatially and temporally distributed content-fixations are themselves precisely locatable in both space and time, but their onsets do *not* mark the onset of awareness of their content. It is always an open question whether any particular content thus discriminated will eventually appear as an element in conscious experience. These distributed content-discriminations yield, over the course of time, something *rather like* a narrative stream or sequence, subject to continual editing by many processes distributed around in the brain, and continuing indefinitely into the future (cf. Calvin's [1990] model of consciousness as "scenario-spinning"). This stream of contents is only rather like a narrative because of its multiplicity; at any point in time there are multiple "drafts" of narrative fragments at various stages of "editing" in various places in the brain. Probing this stream at different intervals produces different effects, elicits different narrative accounts from the subject. If one delays the probe too long (overnight, say) the result is apt to be no narrative left at all – or else a narrative that has been digested or "rationally reconstructed" to the point that it has minimal integrity. If one probes "too early," one may gather data on how early a particular discrimination is achieved in the stream, but at the cost of disrupting the normal progression of the stream. Most important, the Multiple Drafts model avoids the tempting mistake of supposing that there must be a single narrative (the "final" or "published" draft) that is canonical – that represents the *actual* stream of consciousness of the subject, whether or not the experimenter (or even the subject) can gain access to it.

The main points at which this model disagrees with the competing tacit model of the Cartesian Theater, may be summarized:

1. Localized discriminations are *not* precursors of re-*presentations* of the discriminated content for consideration by a more central discriminator.

2. The objective temporal properties of discriminatory states may be determined, but they do *not* determine temporal properties of subjective experience.

3. The "stream of consciousness" is *not* a single, definitive narrative. It is a parallel stream of conflicting and continuously revised contents, no one narrative thread of which can be singled out as canonical – as the true version of conscious experience.

The different implications of these two models will be exhibited by considering several puzzling phenomena that seem at first to indicate that the mind "plays tricks with time." (Other implications of the Multiple Drafts model are examined at length in Dennett 1991b.)

1.2. Some "temporal anomalies" of consciousness. Under various conditions people report experiences in which the temporal ordering of the elements in their consciousness, or the temporal relation of those elements to concurrent activity in their brains, seems to be anomalous or even paradoxical. Some theorists (Libet 1982; 1985a; Popper & Eccles 1977) have argued that these temporal anomalies are proof of the existence of an immaterial mind that interacts with the brain in physically inexplicable fashion. Others (Goodman 1978; Libet 1985b), although eschewing any commitment to dualism, have offered interpretations of the phenomena that seem to defy the accepted temporal sequence of cause and effect. Most recently, another theorist (Penrose 1989 – see also multiple book review in *BBS* 13 (4) 1990) has suggested that a materialistic explanation of these phenomena would require a revolution in fundamental physics. These radical views have been vigorously criticized, but the criticisms have overlooked the possibility that the appearance of anomaly in these cases results from conceptual errors that are so deeply anchored in everyday thinking that even many of the critics have fallen into the same traps. We agree with Libet and others that these temporal anomalies are significant, but we hold a different opinion about what they signify.

We focus on four examples, summarized below. Two, drawn from the work of Libet, have received the most attention and provoked the most radical speculation, but because technical criticisms of his experiments and their interpretation raise doubts about the existence of the phenomena he claims to have discovered, we begin with a discussion of two simpler phenomena whose existence has not been questioned but whose interpretation raises the same fundamental problems. We use these simpler cases to illustrate the superiority of the Multiple Drafts model to the traditional Cartesian Theater model, and then apply the conclusions drawn in the more complicated setting of the controversies surrounding Libet's work. Our argument is that even if Libet's phenomena were not known to exist, theory can readily account for the possibility of phenomena of this pseudo-anomalous sort, and even predict them.

A. Color phi. Many experiments have demonstrated the existence of apparent motion, or the phi phenomenon (Kolers & von Grünau 1976; see also Kolers 1972; van der Waals & Roelofs 1930; and the discussion in Goodman 1978). If two or more small spots separated by as much as 4 degrees of visual angle are briefly lit in rapid succession, a single spot will seem to move. This is the basis of our experience of motion in motion pictures and television. First studied systematically by Wertheimer (1912; for a historical account, see Kolers 1972; Sarris 1989), phi has been subjected to many variations; one of the most striking is reported in Kolers and von Grünau (1976). The philosopher Nelson Goodman had asked Kolers whether the phi phenomenon would persist if the two illuminated spots were different in color, and if so, what would happen to the color of "the" spot as "it" moved? Would the illusion of motion disappear, to be replaced by two separately flashing spots? Would the illusory "moving" spot gradually change from one color to another, tracing a trajectory around the color wheel? The answer, when Kolers and von Grünau performed the experiments, was striking: The spot seems to begin moving and then to change color abruptly *in the middle of its illusory passage* toward the second location. Goodman wondered: "How are we able . . . to fill in the spot at the intervening place-times along a path running from the first to the second flash *before that second flash occurs*?" (1978, p. 73; the same question can be raised about any phi, but the color-switch in midpassage vividly brings out the problem.) Unless there is precognition, the illusory content cannot be created until *after* some identification of the second spot occurs in the brain. But if this identification of the second spot is already "in conscious experience" would it not be too late to interpose the illusory color-switching-while-moving scene between the conscious experience of spot 1 and the conscious experience of spot 2? How does the brain accomplish this sleight-of-hand? Van der Waals and Roelofs (1930) proposed that the intervening motion is produced retrospectively, built only after the second flash occurs, and "projected backwards in time" (Goodman 1978, p. 74), a form of words reminiscent of Libet's "backwards referral in time." But what does it mean that this experienced motion is "projected backwards in time"?

B. The cutaneous "rabbit." The subject's arm rests cushioned on a table, and mechanical square-wave tappers are placed at two or three locations along the arm, up to a foot apart (Geldard & Sherrick 1972; see also Geldard 1977; Geldard & Sherrick 1983; 1986). A series of rhythmical taps is delivered, for example, 5 at the wrist followed by 2 near the elbow and then 3 more on the upper arm. These taps are delivered with interstimulus intervals of between 50 and 200 msec. So a train of taps might last less than a second, or as long as two or three seconds. The astonishing effect is that the taps seem to the subjects to travel in regular sequence over equidistant points up the arm – as if a little animal were hopping along the arm. Now *how did the brain know* that after the 5 taps on the wrist there were going to be some taps near the elbow? The experienced "departure" of the taps from the wrist begins with the second one, yet in catch trials in which the later elbow taps are never delivered, all five wrist taps are felt at the wrist in the expected manner. The brain obviously cannot "know" about a tap at the elbow until after it happens. Perhaps, one might speculate, the brain delays the conscious experience until after all the taps have been "received" and then, somewhere upstream of the seat of consciousness (whatever that is), *revises* the data to fit a theory of motion, and sends the edited version on to consciousness. But would the brain always delay response to one tap in case more came? If not, how does it "know" when to delay?

C. "Referral backwards in time." Since Penfield and Jasper (1954) it has been known that direct electrical stimulation of locations on the somatosensory cortex can induce sensations on corresponding parts of the body. For instance, stimulation of a point on the left somatosensory

cortex can produce the sensation of a brief tingle in the subject's right hand. Libet compared the time course of such cortically induced tingles to similar sensations produced in the more usual way, by applying a brief electrical pulse to the hand itself (Libet 1965; 1981; 1982; 1985a; Libet et al. 1979; see also Churchland 1981a; 1981b; Dennett 1979; Honderich 1984; Popper & Eccles 1977). He argued that although in each case it took considerable time (approximately 500 msec) to achieve "neuronal adequacy" (the stage at which cortical processes culminate to yield a conscious experience of a tingle), when the hand itself was stimulated, the experience was "automatically . . . referred backwards in time."

Most strikingly, Libet reported instances in which a subject's left *cortex* was stimulated *before* his left *hand* was stimulated, something one would tend to expect to give rise to two felt tingles: First right hand (cortically induced) and then left hand. In fact, however, the subjective report was reversed: "first left, then right." Even in cases of simultaneous stimulation, one might have thought, the left-hand tingle should be felt second, because of the additional distance (close to a meter) nerve impulses from the left hand must travel to the brain.

Libet interprets his results as raising a serious challenge to materialism: "A dissociation between the timings of the corresponding 'mental' and 'physical' events would seem to raise serious though not insurmountable difficulties for the . . . theory of psychoneural identity" (1979, p. 222). According to Eccles, this challenge cannot be met:

> This antedating procedure does not seem to be explicable by any neurophysiological process. Presumably it is a strategy that has been learnt by the self-conscious mind . . . the antedating sensory experience is attributable to the ability of the self-conscious mind to make slight temporal adjustments, i.e., to play tricks with time. (Popper & Eccles 1977, p. 364)

D. Subjective delay of consciousness of intention. In other experiments, Libet asked subjects to make "spontaneous" decisions to flex one hand at the wrist while noting the position of a revolving spot (the "second hand" on a clock, in effect) at the precise time they formed the intention (Libet 1985a; 1987; 1989a; see also the accompanying commentaries). Subjects' reports of these subjective simultaneities were then plotted against the timing of relevant electrophysiological events in their brains. Libet found evidence that these "conscious decisions" lagged between 350 and 400 msec behind the onset of "readiness potentials" he was able to record from scalp electrodes, which, he claims, tap the neural events that determine the voluntary actions performed. He concludes that "cerebral initiation of a spontaneous voluntary act begins unconsciously" (1985a, p. 529). That one's consciousness might lag behind the brain processes that control one's body seems to some an unsettling and even depressing prospect, ruling out a real (as opposed to illusory) "executive role" for "the conscious self." (See the discussions by many commentators in *BBS*: Eccles 1985; Mortenson 1985; Van Gulick 1985; and in Pagels 1988, pp. 233ff; and Calvin 1990, pp. 80–81. But see, for a view close to ours, Harnad 1982.)

In none of these cases would there be prima facie evidence of any anomaly were we to forego the opportunity to record the subjects' *verbal reports* of their experiences and subject them to semantic analysis. No sounds appear to issue from heads before lips move, nor do hands move before the brain events that purportedly cause them, nor do events occur in the cortex in advance of the stimuli that are held to be their source. Viewed strictly as the internal and external behavior of a biologically implemented control system for a body, the events observed and clocked in the experiments mentioned exhibit no apparent violations of everyday mechanical causation – of the sort to which Galilean/Newtonian physics provides the standard approximate model. Libet said it first: "It is important to realize that these subjective referrals and corrections are apparently taking place at the level of the *mental* 'sphere'; they are not apparent, as such, in the activities at neural levels" (1982, p. 241).

Put more neutrally (pending clarification of what Libet means by the "mental 'sphere'"), only through the subjects' verbalizations about their subjective experiences do we gain access to a perspective from which the anomalies can appear.[2] Once their verbalizations (including communicative button-pushes, etc.; Dennett 1982), are interpreted as a sequence of speech acts, their *content* yields a time series, *the subjective sequence of the stream of consciousness.* One can then attempt to put this series into registration with another time series, *the objective sequence of observed events in the environment and in the nervous system.* It is the apparent failures of registration, holding constant the assumption that causes precede their effects, that constitute the supposed anomalies (cf. Hoy 1982).

One could, then, "make the problems disappear" by simply refusing to take introspective reports seriously. Although some hearty behaviorists may cling comfortably to the abstemious principle, "Eschew content!" (Dennett 1978), the rest of us prefer to accept the challenge to make sense of what Libet calls "a primary phenomenological aspect of our human existence in relation to brain function" (1985a, p. 534).

> The reports by subjects about their different experiences . . . were not theoretical constructs but empirical observations. . . . The method of introspection may have its limitations, but it can be used appropriately within the framework of natural science, and it is absolutely essential if one is trying to get some experimental data on the mind-brain problem. (Libet 1987, p. 785)

In each example an apparent dislocation in time threatens the *prima facie* plausible thesis that our conscious perceptions are caused by events in our nervous systems, and our conscious acts, in turn, cause events in our nervous systems that control our bodily acts. To first appearances, the anomalous phenomena show that these two standard causal links cannot be sustained unless we abandon a foundational – some would say a logically necessary – principle: *Causes precede their effects.* It seems that in one case (subjective delay of awareness of intention), our conscious intentions *occur too late* to be the causes of their bodily expressions or implementations, and in the other cases, percepts *occur too early* to have been caused by their stimuli. The vertiginous alternative, that something in the brain (or "conscious self") can "play tricks with time" by "projecting" mental events backwards in time, would require us to abandon the foundational principle that causes precede their effects.

There is a widespread conviction that no such revolutionary consequence follows from any of these phenomena, a conviction we share. But some of the influential arguments that have been offered in support of this conviction persist in a commitment to the erroneous presuppositions that made the phenomena appear anomalous in the first place. These presuppositions are all the more insidious because although in their overt, blatant forms they are roundly disowned by one and all, they creep unnoticed back into place, distorting analysis and blinding theory-builders to other explanations.

2. The models in action: Diagnosing the tempting errors

2.1. The representation of temporal properties versus the temporal properties of representations. The brain, as the control system responsible for solving a body's real-time problems of interaction with the environment, is under significant time pressure. It must often arrange to modulate its output in light of its input within a time window that leaves no slack for delays. In fact, many acts can be only *ballistically* initiated; there is no time for feedback to adjust the control signals. Other tasks, such as speech perception, would be beyond the physical limits of the brain's machinery if they did not use ingenious anticipatory strategies that feed on redundancies in the input (Libermann 1970).

How, then, does the brain keep track of the temporal information it manifestly needs? Consider the following problem: Because the toe-brain distance is much greater than the hip-brain distance, or the shoulder-brain distance or the forehead-brain distance, stimuli delivered simultaneously at these different sites will arrive at Headquarters in staggered succession, if travel-speed is constant along all paths. How (one might be tempted to ask) does the brain "ensure central simultaneity of representation for distally simultaneous stimuli"? This encourages one to hypothesize some "delay loop" mechanism that could store the early arrivers until they could be put "in synch" with the latecomers, but this is a mistake. The brain should not solve *this* problem, for an obvious engineering reason: It squanders precious time by committing the full range of operations to a "worst case" schedule. Why should important signals from the forehead (for instance) dawdle in the anteroom just because there might someday be an occasion when concurrent signals from the toes need to be compared to (or "bound to") them?

The brain sometimes uses "buffer memories" to cushion the interface between its internal processes and the asynchronous outside world (Neisser 1967; Newell et al. 1989; Sperling 1960), but there are also ways for the brain to use the temporal information it needs without the delays required for imposing a master synchrony. The basic design principle is well illustrated in an example in which a comparable problem is confronted and (largely) solved, though on a vastly different temporal and spatial scale.

Consider the communication difficulties faced by the far-flung British Empire before the advent of radio and telegraph, as illustrated by the Battle of New Orleans. On January 8, 1815, 15 days after the truce was signed in Belgium, more than a thousand British soldiers were killed in this needless battle. We can use this debacle to see how the system worked. Suppose on Day 1 the treaty is signed in Belgium, with the news sent by land and sea to America, India, Africa. On Day 15 the battle is fought in New Orleans, and news of the defeat is sent by land and sea to England, India, and so on. On Day 20, too late, the news of the treaty (and the order to surrender) arrives in New Orleans. On Day 35, let's suppose, the news of the defeat arrives in Calcutta, but the news of the treaty doesn't arrive there until Day 40 (via a slow overland route). To the commander-in-chief in Calcutta, the battle would "seem" to have been fought before the treaty was signed – were it not for the practice of dating letters, which permits him to make the necessary correction.

These communicators solved their problems of communicating information about time by embedding representations of the relevant time information in the *content* of their signals, so that the arrival time of the signals themselves was *strictly irrelevant* to the information they carried. A date written at the head of a letter (or a dated postmark on the envelope) gives the recipient information about when it was sent, information that survives any delay in arrival.[3] This distinction between time represented (by the postmark) and time of representing (the day the letter arrives) is an instance of a familiar distinction between content and vehicle, and although the details of this particular solution are not available to the brain's communicators (because they don't "know the date" when they send their messages), the general principle of the content/vehicle distinction is relevant to information-processing models of the brain in ways that have not been well appreciated.[4]

In general, we must distinguish features of representings from the features of representeds (Neumann 1990); someone can shout "softly, on tiptoe" at the top of his lungs, there are gigantic pictures of microscopic objects and oil paintings of artists making charcoal sketches. The top sentence of a written description of a standing man need not describe his head, nor the bottom sentence his feet. To suppose otherwise is confusedly to superimpose two different spaces: The representing space and the represented space. The same applies to time. Consider the *spoken phrase*, "a bright, brief flash of red light." The beginning of *it* is "a bright" and the end of it is "red light." Those portions of that speech event are not themselves representations of onsets or terminations of a brief red flash (cf. Efron 1967, p. 714). No informing event in the nervous system can have zero duration (any more than it can have zero spatial extent), so it has an onset and termination separated by some amount of time. If it *represents* an event in experience, then the event it represents must itself have nonzero duration, an onset, a middle, and a termination. But there is no reason to suppose that the beginning of the representing represents the beginning of the represented.[5]

Similarly, the representing by the brain of "A before B" does not have to be accomplished by first:

a representing of A,

followed by:

a representing of B.

"B after A" is an example of a (spoken) vehicle that

represents A as being before B, and the brain can avail itself of the same freedom of temporal placement. What matters for the brain is not necessarily *when* individual representing events happen in various parts of the brain (as long as they happen in time to control the things that need controlling!) but their *temporal content*. That is, what matters is that the brain can proceed to control events "under the assumption that A happened before B" whether or not the information that A has happened enters the relevant system of the brain and gets recognized as such before or after the information that B has happened. (Recall the commander-in-chief in Calcutta: First he is informed of the battle, and then he is informed of the truce, but because he can extract from this the information that the truce came first, he can act accordingly.) Systems in various locations in the brain can, in principle, avail themselves of similar information-processing, and that is why fixing the exact time of onset of some representing element in some place in the brain does not provide a temporal landmark relative to which other elements *in the subjective sequence* can – or must – be placed.

How are temporal properties really inferred by the brain? Systems of "date stamps" or "postmarks" are not theoretically impossible (Glynn 1990), but there is a cheaper, less foolproof but biologically more plausible way: by what we might call *content-sensitive settling*. A useful analogy would be the film studio where the sound track is "synchronized" with the film. The various segments of audio tape may by themselves have lost all their temporal markers, so that there is no simple, mechanical way of putting them into apt registration with the images. But sliding them back and forth relative to the film and looking for convergences, will usually swiftly home in on a "best fit." The slap of the slateboard at the beginning of each take provides a double saliency, an auditory and a visual clap, to slide into synchrony, pulling the rest of the tape and the frames into position at the same time. But there are typically so many points of mutually salient correspondence that this conventional saliency at the beginning of each take is just a handy redundancy. Getting the registration right depends on the *content* of the film and the tape, but not on sophisticated analysis of the content. An editor who knew no Japanese would find synchronizing a Japanese soundtrack to a Japanese film difficult and tedious but not impossible. Moreover, the temporal order of the stages of the process of putting the pieces into registration is independent of the content of the product; the editor can organize scene three before organizing scene two, and in principle could even do the entire job running the segments "in reverse."

Quite "stupid" processes can do similar jiggling and settling in the brain. The computation of depth in random-dot stereograms (Julesz 1971) is a spatial problem for which we can readily envisage temporal analogues. If the system receives stereo pairs of images, the globally optimal registration can be found without first having to subject each data array to an elaborate process of feature extraction. There are enough lowest-level coincidences of saliency – the individual dots in a random dot stereogram – to dictate a solution. In principle, then, the brain can solve some of its problems of temporal inference by such a process, drawing data not from left and right eyes, but from whatever information-sources are involved in a process requiring temporal judgments. (See Gallistel, 1990, especially pp. 539–49, for a discussion of the requirements for "spatiotemporal specification.")

Two important points follow from this. First, such temporal inferences can be drawn (such temporal discriminations can be made) by comparing the (low-level) *content* of several data arrays, and this real time process need not occur in the temporal order that its product eventually represents. Second, once such a temporal inference has been drawn, which may be *before* high-level features have been extracted by other processes, it does not have to be drawn again! There does not have to be a *later* representation in which the high-level features are "presented" in a real time sequence for the benefit of a second sequence-judger. In other words, having drawn inferences from these juxtapositions of temporal information, the brain can go on to represent the results in any format that fits its needs and resources – not necessarily a format in which "time is used to represent time."

There remains a nagging suspicion that whereas the brain may take advantage of this representational freedom for other properties, it cannot do so for the property of temporal sequence. Mellor explicitly enunciates this assumption, deeming it too obvious to need support:

> Suppose for example I see one event *e* precede another, *e**. I must first see *e* and then *e**, my seeing of *e* being somehow recollected in my seeing of *e**. That is, my seeing of *e* affects my seeing of *e**: This is what makes me – rightly or wrongly – see *e* precede *e**rather than the other way round. But seeing *e* precede *e** means seeing *e* first. So the causal order of my perceptions of these events, by fixing the temporal order I perceive them to have, fixes the temporal order of the perceptions themselves. . . . the striking fact . . . should be noticed, namely that perceptions of temporal order need temporally ordered perceptions. *No other property or relation has to be thus embodied in perceptions of it* [our emphasis]: perceptions of shape and colour, for example, need not themselves be correspondingly shaped or coloured. (Mellor 1981, p. 8)

We believe this is false, but there is something right about it. Because the fundamental function of representation in the brain is to control behavior in real time, the timing of representings is *to some degree* essential to their task, in two ways. First, the timing may, at the outset of a perceptual process, be *what determines the content*. Consider how to distinguish a spot moving from right to left from a spot moving from left to right on a motion picture screen. The *only* difference between the two may be the temporal order in which two frames (or more) are projected. If the brain determines "first A, then B" the spot is seen as moving in one direction; if the brain determines "first B, then A" the spot is seen as moving in the opposite direction. This discrimination is, then, as a matter of logic, based on the brain's capacity to make a temporal order judgment of a particular level of resolution. Motion picture frames are usually exposed at the rate of 24 per second, and so the visual system can resolve order between stimuli that occur within about 50 msec. This means that the actual temporal properties of signals – their onset times, their velocity in the system, and hence their arrival times – must be accurately controlled until such a discrimination is made. But once it is made locally by some circuit in the visual system (even as peripherally

as the ganglion cells of the rabbit's retina! – Barlow & Levick 1965), the content "from left to right" can then be sent, in a temporally sloppy way, anywhere in the brain where this directional information might be put to use. This way one can explain the otherwise puzzling fact that at interstimulus intervals at which people are unable to perform above chance on temporal order judgments, they perform flawlessly on other judgments that logically call for the same temporal acuity. Thus Efron (1973) showed that subjects could easily distinguish sounds, flashes, and vibrations that differed only in the order in which two component stimuli occurred at a fraction of the interstimulus interval at which they can explicitly specify their order.

A second constraint on timing has already been noted parenthetically above: It does not matter in what order representations occur so long as they occur in time to contribute to the control of the appropriate behavior. The function of a representing may depend on meeting a *deadline*, which is a temporal property of the vehicle doing the represent*ing*. This is particularly evident in such time-pressured environments as the imagined Strategic Defense Initiative. The problem is not how to make computer systems represent, accurately, missile launches, but how to represent a missile launch accurately during the brief time while one can still do something about it. A message that a missile was launched at 6:04:23.678 A.M. EST may accurately represent the time of launch forever, but its utility may utterly lapse at 6:05 A.M. EST. For any task of control, then, there is a *temporal control window* within which the temporal parameters of representings may in principle be moved around ad lib.

The deadlines that limit such windows are not fixed, but rather depend on the task. If, rather than intercepting missiles, you are writing your memoirs or answering questions at the Watergate hearings (Neisser 1981), you can recover the information you need about the sequence of events in your life to control your actions in almost any order, and you can take your time drawing inferences.

These two factors explain what is plausible in Mellor's claim, without supporting the invited conclusion that all perceptions of temporal order must be accomplished in a single place by a process that observes *seriatim* a succession of "perceptions" or other representations. Once the perceptual processes *within* an observer have begun to do their work, providing the necessary discriminations, there is no point in undoing their work to provide a job for a yet more interior observer.

Causes must precede effects. This fundamental principle ensures that temporal control windows are bounded at both ends: by the earliest time at which information could arrive in the system, and by the latest time at which information could contribute causally to the control of a particular behavior. Moreover, the principle applies to the multiple distributed processes that achieve such control. Any particular process that requires information from some source must indeed wait for that information; it can't get there till it gets there. This is what rules out "magical" or precognitive explanations of the color-switching phi phenomenon, for example. The content *green spot* cannot be attributed to any event, conscious or unconscious, until the light from the green spot has reached the eye and triggered the normal neural activity in the visual system up to the level at which the discrimination of green is accomplished. Moreover, all content reported or otherwise expressed in subsequent behavior must have been "present" (in the relevant place in the brain, but not necessarily in consciousness) in time to have contributed causally to that behavior. For instance, if a subject in an experiment *says* "dog" in response to a visual stimulus, we can work backwards from the behavior, which was clearly controlled by a process that had the content *dog* (unless the subject says "dog" to every stimulus, or spends the day saying "dog dog dog . . . " etc.) And since it takes on the order of 100 msec to execute a speech intention of this sort, we can be quite sure that the content *dog* was present in (roughly) the language areas of the brain by 100 msec before the utterance. Working from the other end, we can determine the earliest time the content *dog* could have been computed or extracted by the visual system from the retinal input, and even, perhaps, follow its creation and subsequent trajectory through the visual system and into the language areas.

What would be truly anomalous (indeed a cause for lamentations and the gnashing of teeth) would be if the time that elapsed between the *dog*-stimulus and the "dog"-utterance were less than the time physically required for this content to be established and moved through the system. No such anomalies have been uncovered, however. It is only when we try to put the sequence of events thus detectable in the objective processing stream into registration with the subject's subjective sequence *as indicated by what the subject subsequently says* that we have any sign of anomaly at all.

2.2. Orwellian and Stalinesque revisions: The illusion of a distinction. Now let us see how the two different models, the Cartesian Theater and Multiple Drafts, deal with the presumed anomalies, starting with the simpler and less controversial phenomena. The Cartesian Theater model postulates a place within the brain where what happens "counts"; that is, it postulates that the features of events occurring within this functionally definable boundary (whatever it is) are definitive or constitutive features of conscious experience. (The model applies to all features of subjective experience, but we are concentrating on temporal features.) This implies that all revisions of content accomplished by the brain can be located relative to this place, a deeply intuitive – but false – implication that can be illustrated with a thought experiment.

Suppose we tamper with your brain, inserting in your memory a bogus woman wearing a hat where none was (e.g., at the party on Sunday). If on Monday, when you recall the party, you remember her, and can find no internal resources for so much as doubting the veracity of your memory, we could all agree that you never *did* experience her; that is, not at the party on Sunday. Of course your subsequent experience of (bogus) recollection can be as vivid as may be, and on Tuesday we can certainly agree that you have had vivid conscious experiences of there being a woman in a hat at the party, but the *first* such experience, we would insist, was on Monday, not Sunday (although it doesn't seem this way to you).

We lack the power to insert bogus memories by neurosurgery, but sometimes our memories play tricks on us, so what we cannot yet achieve surgically happens in the

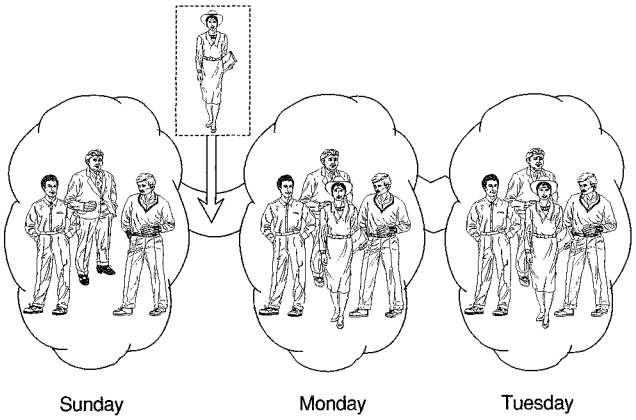

Figure 1. Post-experiential memory tampering.

brain on its own. Sometimes we seem to remember, even vividly, experiences that never occurred. We might call such post-experiential contaminations or revisions of memory *Orwellian*, recalling George Orwell's chilling vision of the Ministry of Truth in *1984*, which busily rewrote history and thus denied access to the (real) past to all who followed.

Orwellian revision is one way to fool posterity. Another is to stage show trials, carefully scripted presentations of false testimony and bogus confessions, complete with simulated evidence. We might call this ploy *Stalinesque*. Notice that if we are usually sure which mode of falsification has been attempted on us, the Orwellian or the Stalinesque, this is just a happy accident. In any *successful* disinformation campaign, were we to wonder whether the accounts in the newspapers were Orwellian accounts of trials that never happened at all, or true accounts of phony show trials that actually did happen, we might be unable to tell the difference. If *all* the traces – newspapers, videotapes, personal memoirs, inscriptions on gravestones, living witnesses, and so on – have been either obliterated or revised, we will have no way of knowing which sort of fabrication happened: a fabrication *first*, culminating in a staged trial whose accurate history we now have before us, or *after* a summary execution, history-fabrication covering up the deed. No trial of any sort *actually* took place.

The distinction between reality and (subsequent) appearance, and the distinction between Orwellian and Stalinesque methods of producing misleading archives, work unproblematically in the everyday world, at macroscopic time scales. One might well think these distinctions apply unproblematically *all the way in*. That is the habit of thought that produces the cognitive illusion of Cartesian materialism. We can catch it in the act in a thought experiment that differs from the first one in nothing but time scale.

Suppose a long-haired woman jogs by. About one second *after* this, a subterranean memory of some earlier woman – a short-haired woman with glasses – contaminates the memory of what you have just seen: When asked a minute later for details of the woman you just saw, you report, sincerely but erroneously, that she was wearing glasses. Just as in the previous case, we are inclined to say that your original *visual* experience, as opposed to the memory of it seconds later, was *not* of a woman with

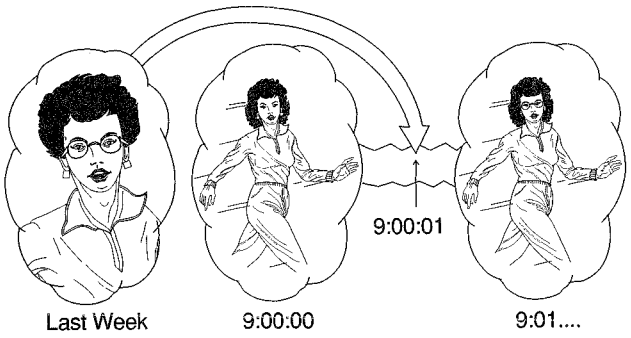

Figure 2. Orwellian revision.

glasses. But because of the subsequent memory-contaminations, it seems to you exactly as if at the first moment you saw her, you were struck by her eyeglasses. An Orwellian, postexperiential revision has happened: There was a fleeting instant, before the memory contamination took place, when it *didn't* seem to you she had glasses. For that brief moment, the *reality* of your conscious experience was a long-haired woman *without* eyeglasses, but this historical fact has become inert; it has left no trace, thanks to the contamination of memory that came one second after you glimpsed her.

This understanding of what happened is jeopardized by an alternative account, however. Your subterranean earlier memories of that short-haired woman with the glasses could just as easily have contaminated your experience *on the upward path*, in the processing of information that occurs "prior to consciousness" so that you actually *hallucinated* the eyeglasses from the very beginning of your experience. In that case, your obsessive memory of the woman with glasses would be playing a Stalinesque trick on you, creating a "show trial" for you to experience, which you then accurately recall at later times, thanks to the record in your memory. To naive intuition these two cases are as different as can be. Told the first way (Figure 2), you suffer no hallucination at the time the woman jogs by, but suffer subsequent memory-hallucinations: You have false memories of your actual ("real") experience. Told the second way (Figure 3), you hallucinate when she runs by, and then accurately remember that hallucination (which "really did happen in consciousness") thereafter. Surely these are distinct possibilities, no matter how finely we divide up time?

No. Here the distinction between perceptual revisions and memory revisions that works so crisply at other scales is not guaranteed application. We have moved into the

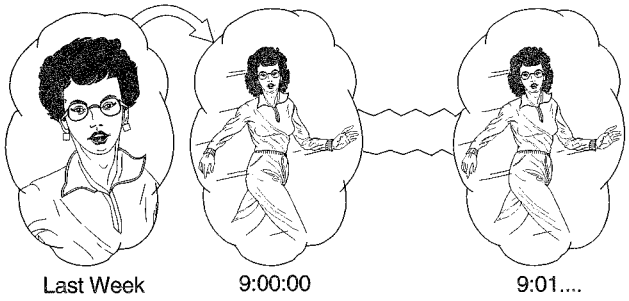

Figure 3. Stalinesque show trial.

foggy area in which the subject's point of view is spatially and temporally smeared, and the question *Orwellian or Stalinesque?* (post-experiential or pre-experiential) need have no answer. The boundary between perception and memory, like most boundaries between categories, is not perfectly sharp, as has often been noted.

There is a time window that began when the long-haired woman jogged by, exciting your retinas, and ended when you expressed – to yourself or someone else – your eventual conviction that she was wearing glasses. At some time during this interval, the content *wearing glasses* was spuriously added to the content *long-haired woman*. We may assume (and might eventually confirm in detail) that there was a brief time when the content *long-haired woman* had already been discriminated in the brain but *before* the content *wearing glasses* had been erroneously "bound" to it. Indeed, it would be plausible to suppose that this discrimination of a long-haired woman was what triggered the memory of the earlier woman with the glasses. What we would not know, however, is whether this spurious binding was before or after the fact – the presumed fact of "actual conscious experience." Were you first conscious of a long-haired woman without glasses and then conscious of a long-haired woman with glasses, a subsequent consciousness that wiped out the memory of the earlier experience, or was the very first instant of conscious experience already spuriously tinged with eyeglasses? If Cartesian materialism were correct, this question would have to have an answer, even if we – and you – could not determine it retrospectively by any test, for the content that "crossed the finish first" was either *long-haired woman* or *long-haired woman with glasses*. But what happens to this question if Cartesian materialism is incorrect (as just about everyone agrees)? Can the distinction between pre-experiential and post-experiential content revisions be maintained?

An examination of the color phi phenomenon shows that it cannot. On the first trial (i.e., without conditioning), subjects *report* seeing the color of the moving spot switch in midtrajectory from red to green – a report sharpened by Kolers's ingenious use of a pointer device which subjects retrospectively-but-as-soon-as-possible "superimposed" on the trajectory of the illusory moving spot; such pointer locations had the content: "The spot changed color right about *here*" (Kolers & von Grünau 1976, p. 330). Recall Goodman's (1978, p. 73) expression of the puzzle: "How are we able . . . to fill in the spot at the intervening place-times along a path running from the first to the second flash *before that second flash occurs*?"

Consider, first, a Stalinesque mechanism: In the brain's editing room, located before consciousness, there is a delay, a loop of slack like the "tape delay" used in broadcasts of "live" programs, which gives the censors in the control room a few seconds to bleep out obscenities before broadcasting the signal. *In the editing room*, first frame A, of the red spot, arrives, and then, when frame B, of the green spot, arrives, some interstitial frames (C and D) can be created and then spliced into the film (in the order A, C, D, B) on its way to projection in the theater of consciousness. By the time the "finished product" arrives at consciousness, it already has its illusory insertion.

Alternatively, there is the hypothesis of an Orwellian mechanism: Shortly after the awareness of the first spot *and* the second spot (with no illusion of apparent motion at all), a revisionist historian of sorts, in the brain's memory-library receiving station, notices that the unvarnished history of this incident doesn't make enough sense, so he "interprets" the brute events, red-followed-by-green, by making up a narrative about the intervening passage, complete with midcourse color change, and installs this history, incorporating his glosses, frames C and D (in Figure 4), in the memory library for all future reference. Because he works fast, within a fraction of a second – the amount of time it takes to frame (but not utter) a verbal report of what you have experienced – the record you rely on, stored in the library of memory, is already contaminated. You *say* and *believe* that you saw the illusory motion and color change, but that is really a memory hallucination, not an accurate recollection of your original awareness.

How could we see which of these hypotheses is correct? It might seem that we could rule out the Stalinesque hypothesis quite simply, because of the delay in consciousness it postulates. In Kolers and von Grünau's experiment, there was a 200 msec difference in onset between the red and green spot, and since, *ex hypothesi*, the *whole experience* cannot be composed by the editing room until after the content *green spot* has reached the editing room, consciousness of the initial red spot will have to be delayed by at least that much. (If the editing room sent the content *red spot* up to the theater of consciousness immediately, before receiving frame B and then fabricating frames C and D, the subject would presumably experience a gap in the film, a noticeable delay of around 200 msec between A and C.)

Suppose we ask subjects to press a button "as soon as you experience a red spot." We would find little or no difference in response time to a red spot alone versus a red spot followed 200 msec later by a green spot (in which case the subjects report color-switching apparent motion). This could be because there is *always* a delay of at least 200 msec in consciousness, but aside from the biological implausibility of such a squandering of time, there is the evidence from many quarters that responses under conscious control, although slower than such responses as reflex blinks, occur with close to the minimum latencies that are physically possible; after subtracting the demonstrable travel times for incoming and outgoing pulse trains, and the response preparation time, there is little time left over in "central processing" in which to hide a 200 msec delay. So the responses had to have been

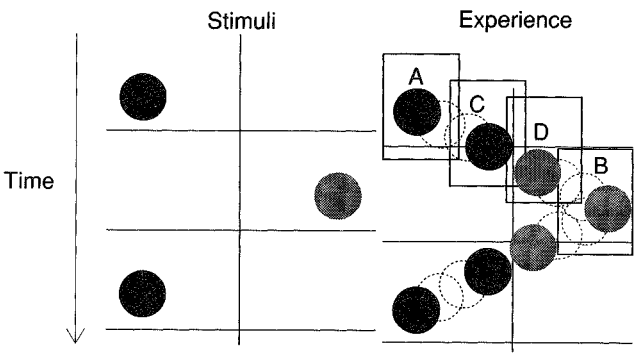

Figure 4. Frames C and D inserted in the editing room.

initiated before the discrimination of the second stimulus, the green spot. This would seem overwhelmingly to favor the Orwellian, post-experiential mechanism: As soon as the subject *becomes conscious* of the red spot, he initiates a button-press. *While that button press is forming*, he becomes conscious of the green spot. *Then* both these experiences are wiped from memory, replaced in memory by the revisionist record of the red spot moving over and then turning green halfway across. He readily and sincerely (but mistakenly) reports having seen the red spot moving toward the green spot before changing color.

If the subject were to insist that he really was conscious from the very beginning of the red spot moving and changing color, the Orwellian theorist would firmly explain to him that he is wrong; his memory is playing tricks on him; the fact that he pressed the button when he did is conclusive evidence that he was conscious of the (stationary) red spot before the green spot had even occurred. After all, his instructions were to press the button *when he was conscious of* a red spot. He must have been conscious of the red spot about 200 msec before he could have been conscious of it moving and turning green. If that is not how it seems to him, he is simply mistaken.

The defender of the Stalinesque (pre-experiential) alternative is not defeated by this, however. Actually, he insists, the subject responded to the red spot *before* he was conscious of it! The directions to the subject (to respond to a red spot) had somehow trickled down from consciousness into the editing room, which *unconsciously* initiated the button-push before sending the edited version (frames ACDB) up to consciousness for "viewing." The subject's memory has played no tricks on him; he is reporting exactly what he was conscious of, unless he insists that he pushed the button after consciously seeing the red spot; his "premature" button-push was unconsciously (or preconsciously) triggered (cf. Velmans 1991).

Where the Stalinesque theory postulates a button-pushing reaction to an *un*conscious detection of a red spot, the Orwellian theory postulates a *conscious* experience of a red spot that is immediately obliterated from memory by its sequel. So here is the rub: We have two different models of what happens in the phi phenomenon: one posits a Stalinesque "filling in" on the upward, pre-experiential path, and the other posits an Orwellian "memory revision" on the downward, post-experiential path, and *both* of them are consistent with *whatever* the subject says or thinks or remembers. Note that the inability to distinguish these two possibilities does not apply only to the *outside observers* who might be supposed to lack some private data to which the subject had "privileged access." You, as a subject in a phi phenomenon experiment, *could not* discover anything in the experience from your own first-person perspective that would favor one theory over the other; the experience would "feel the same" on either account. As the interstimulus interval is lengthened subjects pass from seeing apparent motion to seeing individual stationary flashes. There is an intermediate range of intervals where the phenomenology is somewhat paradoxical: You see the spots as two stationary flashes *and* as one thing moving. This sort of apparent motion is readily distinguishable from the swifter, smoother sort of apparent motion of cinema, for instance, but your capacity to make *this* discrimination is not relevant to the dispute between the Orwellian and the Stalinesque theorist. They agree that you can make this discrimination under the right conditions; what they disagree about is how to describe the cases of apparent motion that you *can't* tell from real motion – the cases in which you really (mis-)*perceive* the illusory motion. To put it loosely, in these cases is your memory playing tricks with you, or are just your eyes playing tricks with you? You can't tell "from the inside."

We can see the same indistinguishability even more clearly when we see how the two different models handle the well-studied phenomenon of *metacontrast* (for a review, see Breitmeyer 1984). If a stimulus is flashed briefly on a screen and then followed, after a brief interstimulus interval, by a second "masking" stimulus, subjects *report* seeing only the second stimulus. (And if you put yourself in the subject's place you will see for yourself; you will be prepared to swear that there was only one flash.) The standard description of such phenomena is that the second stimulus somehow *prevents conscious experience* of the first stimulus (in other words, it somehow waylays the first stimulus on its way to consciousness). But people can nevertheless do much better than chance if required to guess whether there were two stimuli. This only shows once again that stimuli can have their effects on us without our being conscious of them. This standard line is, in effect, the Stalinesque model of metacontrast: The first stimulus never gets to play on the stage of consciousness; it has whatever effects it has entirely unconsciously. But we have just uncovered a second, Orwellian model of metacontrast: Subjects are indeed conscious of the first stimulus (which would "explain" their capacity to guess correctly) but their memory of this conscious experience is almost entirely obliterated by the second stimulus (which is why they deny having seen it, in spite of their tell-tale better-than-chance guesses).[6]

Both the Orwellian and the Stalinesque version of the Cartesian Theater model can deftly account for *all* the data – not just the data we already have, but the data we can imagine getting in the future. They both account for the verbal reports: One theory says they are innocently mistaken whereas the other says they are accurate reports of experienced "mistakes." (A similar verdict is suggested in the commentaries of Holender 1986; see especially Dixon 1986; Erdelyi 1986; Marcel 1986; Merikle & Cheesman 1986.) They agree about just where in the brain the mistaken content enters the causal pathways; they just disagree about whether that location is pre-experiential or post-experiential. They both account for the nonverbal effects: One says they are the result of unconsciously discriminated contents while the other says they are the result of consciously discriminated but forgotten contents. They agree about just where and how in the brain these discriminations occur; they just disagree about whether to interpret those processes as happening inside or outside the charmed circle of consciousness. Finally, they both account for the subjective data – whatever is obtainable "from the first-person-perspective" – because they agree about how it ought to "feel" to subjects: Subjects should be unable to tell the difference between misbegotten experiences and immediately misremembered experiences. So, in spite of first appearances, there is really only a verbal difference between the two theories (cf. Reingold & Merikle 1990). They tell exactly the same story except for where they

place a mythical Great Divide, a point in time (and hence a place in space) whose *fine-grained* location is nothing that subjects can help them locate, and whose location is also neutral with regard to all other features of their theories. This is a difference that makes no difference.

Consider a contemporary analogy. With the advent of word-processing and desktop publishing and electronic mail, we are losing the previously quite hard-edged distinction between pre-publication editing, and post-publication correction of "errata." With multiple drafts in electronic circulation, and with the author readily making revisions in response to comments received by electronic mail, calling one of the drafts the canonical text – the text of "record," the one to cite in one's own publications – becomes a somewhat arbitrary matter. Often most of the intended readers, the readers whose reading of the text matters, read only an early draft; the "published" version is archival and inert. If it is important effects we are looking for, then, most if not all the important effects of writing a text are now spread out over many drafts, not postponed until after publication. It used to be otherwise; virtually all of a text's important effects happened *after* appearance in a book or journal and *because of* its making such an appearance. All the facts are in, and now that the various candidates for the "gate" of publication can be seen no longer to be functionally important, if we feel we need the distinction at all, we will have to decide arbitrarily what is to count as publishing a text. There is no natural summit or turning point in the path from draft to archive.

Similarly – and this is the fundamental implication of the Multiple Drafts model – if one wants to settle on some moment of processing in the brain as the moment of consciousness, this has to be arbitrary. One can always "draw a line" in the stream of processing in the brain, but there are no functional differences that could motivate declaring all prior stages and revisions unconscious or preconscious adjustments and all subsequent emendations to the content (as revealed by recollection) to be post-experiential memory-contamination. The distinction lapses at close quarters.

Another implication of the Multiple Drafts model, in contrast to the Cartesian Theater, is that there is no need – or room – for the sort of "filling in" suggested by frames C and D of Figure 4. Discussing Kolers' experiment, Goodman notes that it "seems to leave us a choice between a retrospective construction theory and a belief in clairvoyance" (1978, p. 83). What then is "retrospective construction"?

> Whether perception of the first flash is thought to be *delayed or preserved or remembered* [our emphasis], I call this the retrospective construction theory – the theory that the construction perceived as occurring between the two flashes is accomplished not earlier than the second.

It seems at first that Goodman does not choose between a Stalinesque theory (perception of the first flash is delayed) and an Orwellian theory (the perception of the first flash is preserved or remembered), but his Orwellian revisionist does not merely adjust judgments; he *constructs* material to *fill in* the gaps: "Each of the intervening places along a path between the two flashes is filled in . . . with one of the flashed colors rather than with successive intermediate colors" (Goodman 1978, p. 85). What Goodman overlooks is the possibility that the brain doesn't actually have to go to the trouble of "filling in" anything with "construction," for no one is looking. As the Multiple Drafts model makes explicit, once a discrimination has been made once, it does not have to be made again; the brain just adjusts to the conclusion that is drawn, making the new interpretation of the information available for the modulation of subsequent behavior. Recall the commander-in-chief in Calcutta; he just had to *judge* that the truce came before the battle; he didn't also have to mount some sort of pageant of "historical reconstruction" to watch, in which he receives the letters in the "proper" order.

Similarly, when Goodman (1978) proposes that "the intervening motion is produced retrospectively, built only after the second flash occurs, and projected backwards in time," this suggests ominously that a final film is made and then run through a magical projector whose beam somehow travels backwards in time onto the mind's screen. Whether or not this is just what Van der Waals and Roelofs (1930) had in mind when they proposed "retrospective construction," it is presumably what led Kolers (1972, p. 184) to reject their hypothesis, insisting that all construction is carried out in "real time." Why, though, should the brain bother to "produce" the "intervening motion"? Why not just conclude that there was intervening motion, and encode that "retrospective" content into the processing stream? This would suffice for it to seem to the subject that intervening motion had been experienced.

Our Multiple Drafts model agrees with Goodman that retrospectively the brain creates the content (the judgment) that there was intervening motion, and this content is then available to govern activity and leave its mark on memory. But our model claims that the brain does not bother "constructing" any representations that go to the trouble of "filling in" the blanks. That would be a waste of time and (shall we say?) *paint*. The judgment is *already in*, so the brain can get on with other tasks![7]

Goodman's "projection backwards in time," like Libet's "backwards referral in time," is an equivocal phrase. It might mean something modest and defensible: A *reference to some past time* is included in the content. On this reading it could be a claim like, "This novel takes us back to ancient Rome," which almost no one would interpret in a metaphysically extravagant way, as claiming that the novel was some sort of time travel machine. This is the reading that is consistent with Goodman's other views, but Kolers apparently took it to mean something metaphysically radical: that there was some actual projection of one thing at one time to another time. As we shall see, the same equivocation bedevils Libet's interpretation of his phenomena.

The model of the Cartesian Theater creates artifactual puzzle questions that cannot be answered, whereas for our model these questions cannot meaningfully arise. This can be seen by applying both models to other experiments that probe the limits of the distinction between perception and memory. A normally sufficient, but not necessary, condition for having experienced something is subsequent verbal report, and this is the anchoring case around which all the puzzle cases revolve. Suppose that although one's brain has registered – that is, responded to – (some aspects of) an event, something intervenes between that internal response and a subse-

quent occasion for verbal report. If there was no time or opportunity for an initial overt response of any sort, and if the intervening events prevent later overt responses (verbal or otherwise) from incorporating reference to some aspect(s) of the first event, this creates a puzzle question: Were they never consciously perceived, or have they been rapidly forgotten?

Consider the familiar span of apprehension. Multiple letters are simultaneously briefly exposed. Some are identified. The rest were certainly seen. The subject insists they were there, knows their number, and has the impression that they were clearcut and distinct. Yet he cannot identify them. Has he failed "really" to perceive them, or has he rapidly "forgotten" them? Or consider an acoustic memory span test, administered at a rapid rate, for example, 4 items a second, so that the subject perforce cannot respond till the acoustic event is over. He identifies some, not others. Yet, subjectively he heard all of them clearly and equally well. Did he not genuinely perceive or did he forget the rest?

And if, under still more constricted circumstances such as metacontrast, the subject even lacks all conviction that the unrecallable items *were there*, should we take this judgment as conclusive grounds for saying he did not experience them, even if they prove to have left other contentful traces on his subsequent behavior? If there is a Cartesian Theater, these questions demand answers, because what gets into the theater and when is supposedly determinate, even if the boundaries appear fuzzy because of human limitations of perception and memory.

Our Multiple Drafts model suggests a different perspective on these phenomena. When a lot happens in a short time, the brain may make simplifying assumptions (for a supporting view, see Marcel 1983). In metacontrast, the first stimulus may be a disc and the second stimulus a ring that fits closely outside the space where the disc was displayed. The outer contour of a disc rapidly turns into the inner contour of a ring. The brain, initially informed just that something happened (something with a circular contour in a particular place), swiftly receives confirmation that there was indeed a ring, with an inner and outer contour. Without further supporting evidence that there was a disc, the brain arrives at the conservative conclusion that there was only a ring. Should we insist that the disc was experienced because *if the ring hadn't intervened* the disc would have been reported? Our model of how the phenomenon is caused shows that there is no motivated way of settling such border disputes: Information about the disc was briefly in a functional position to contribute to a later report, but this state lapsed; there is no reason to insist that this state was inside the charmed circle of consciousness until it got overwritten, or contrarily, to insist that it never quite achieved this state. Nothing discernible to "inside" or "outside" observers could distinguish these possibilities.

In color phi, the processes that calculate that the second spot is green and that there is motion proceed roughly simultaneously (in different parts of the brain) and eventually contribute to the process that concludes that the red spot moved over and abruptly turned green on the way. That conclusion is achieved swiftly enough, in the standard case, to overwhelm or replace any competing contents before they can contribute to the framing of a report. So the subject says – and believes – just what Kolers and von Grünau report, *and that is what the subject was conscious of.* Was the subject *also* conscious a fraction of a second earlier of the stationary red spot? Ask him. If the interstimulus interval is made somewhat longer, there will come a point where the subject *does* report an experience of first a stationary red spot, then a green spot, and then a *noticeably retrospective* sense that the red spot ("must have") moved over and changed color. This experience has – as the subject will tell you – a quite different phenomenology. Apparent motion is experienced under such conditions, but it is obviously different from ordinary motion, and from swifter varieties of apparent motion. How is it different? The subject notices the difference! In this case it does seem to him as if he only later "realized" that there had been motion. But in cases in which this retrospective element is lacking it is still the case that the discrimination of motion-with-color-change is achieved after the colors and locations of the spots were discriminated – and there is no later process of "filling in" required.

In the cutaneous "rabbit," the shift in space (along the arm) is recorded over time by the brain. The number of taps is also recorded. Although in physical reality the taps were clustered at particular locations, the simplifying assumption is that they were distributed regularly across the space-time extent of the experience. The brain relaxes into this parsimonious though mistaken interpretation *after* the taps are registered, and this has the effect of wiping out earlier (partial) interpretations of the taps, but some side effects of those interpretations (e.g., the interpretation that there were five taps, that there were more than two taps, etc.) may live on.

Although different attributes are indeed extracted by different neural facilities at different rates (e.g., location vs. shape vs. color), and although if asked to respond to the presence of each one in isolation we would do so with different latencies, we perceive events, not a successively analyzed trickle of perceptual elements or attributes. As Efron remarks:

> There are no grounds for an a priori assumption that the specificity of our awareness of an object of perception, or an aspect of that object, gradually increases or grows following the moment of its onset from the least specific experience to some maximally specific experience. . . . We do not, when first observing an object with central vision, fleetingly experience the object as it would appear with the most peripheral vision, then as it would appear with less peripheral vision. . . . Similarly, when we shift our attention from one object of awareness to another, there is no experience of "growing" specificity of the new object of awareness – we just perceive the new object. (1967, p. 721)

Is there an "optimal time of probing"? On the plausible assumption that after a while such narratives degrade rather steadily through both fading of details and self-serving embellishment (what I ought to have said at the party tends to turn into what I did say at the party), one can justify probing "as soon as possible" after the stimulus sequence of interest. At the same time, one wants to avoid interfering with the phenomenon by a premature probe. Because perception turns imperceptibly into memory, and "immediate" interpretation turns imperceptibly into rational reconstruction, there is no single, all-context summit on which to direct one's probes. Any probe may

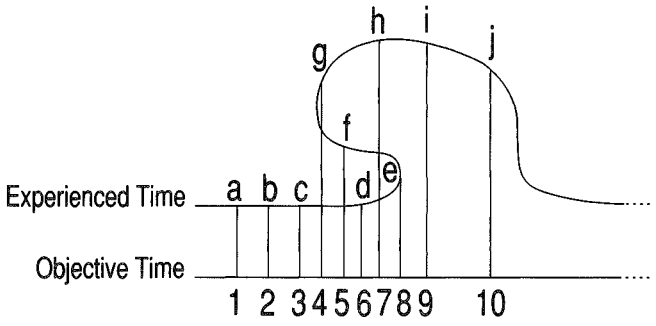

Figure 5. Superimposition of subjective and objective sequences.

elicit a narrative (or narrative fragment), and any such elicited narrative determines a "time line," a subjective sequence of events from the point of view of an observer. This time line may then be compared with other time lines, in particular with the objective sequence of events occurring in the brain of that observer. For the reasons discussed, these two time lines may not superimpose themselves in orthogonal registration. There may be order differences that induce kinks.

There is nothing metaphysically extravagant or challenging about this failure of registration (Snyder 1988). It is no more mysterious or contracausal than the realization that the individual scenes in movies are often shot out of sequence, or that when you read the sentence, "Bill arrived at the party after Sally, but Jane came earlier than either of them," you learn of Bill's arrival before you learn of Jane's earlier arrival. The space and time of the representing is one frame of reference; the space and time of what the representing represents is another. But this metaphysically innocuous fact does nevertheless ground a fundamental metaphysical category: When a portion of the world comes in this way to compose a skein of narratives, that portion of the world is an observer. That is what it is for there to be an observer in the world, a something it is like something to be.

3. The Libet controversies re-examined

3.1. Libet's experiments allegedly showing "backwards referral." Libet's experiments with direct cortical stimulation have provoked a great deal of discussion and speculation, in spite of the fact that they involved very few subjects, were inadequately controlled, and have not been replicated (Churchland 1981a; 1981b). No doubt they have attracted this unusual attention, in spite of their serious technical flaws because, according to Libet, they demonstrate "two remarkable temporal factors":

1. *There is a substantial delay before cerebral activities*, initiated by a sensory stimulus, *achieve "neuronal adequacy"* for eliciting any resulting conscious sensory experience.

2. After neuronal adequacy is achieved, the *subjective timing* of the experience *is (automatically) referred backwards in time*, utilizing a "timing signal" in the form of the initial response of cerebral cortex to the sensory stimulus (1981a, p. 182).

The "timing signal" is the primary evoked potential in the cortex 10 to 20 msec after peripheral stimulation. Libet suggests that the backwards referral is always "to" the timing signal.

Libet's model is Stalinesque: various editing processes occur prior to the moment of "neuronal adequacy," at which time a finished film is projected. How is it projected? Here Libet's account vacillates between an extreme view and a moderate view (cf. Honderich 1984):

a. *Backwards projection*: It is projected backwards in time to some Cartesian Theater where it actually runs in synch with the primary evoked potentials. (The primary evoked potentials, as "timing signals," serve rather like the slateboard used in film-making, showing the projector exactly how far back in time to project the experience.)

b. *Backwards referral*: It is projected in ordinary time, but it carries something like a postmark, reminding the viewer that these events must be understood to have occurred somewhat earlier. (In this case the primary evoked potentials serve simply as dates, which might be *represented* on the Cartesian screen by a title, "On the eve of the Battle of Waterloo" or "New York City, Summer, 1942.")

Libet's own term is "referral" and he defends it by reminding us of the "long recognized and accepted" phenomenon of spatial referral, which might suggest the moderate reading. But because he also insists that this backwards referral is "remarkable" and a challenge to the theory of "psychoneural identity," he invites the extreme interpretation.[8] And his interpretation is further supported by a passage at the close of Libet 1981:

> There is experimental evidence for the view that the subjective or mental "sphere" could indeed "fill in" spatial and temporal gaps. How else, for example, could one view that already mentioned enormous discrepancy *that is known to exist* between a subjective visual image and the configuration of neuronal activities that gives rise to the experience of the image? (p. 196)[9]

Let us consider the details. "Neuronal adequacy," which Libet estimates to require up to 500 msec of cortical activity, is determined by seeing how late, following initial stimulation, a direct cortical stimulation can interfere with the consciousness subsequently reported. Beyond that critical interval, a direct cortical stimulus would be reported by the subject to be a *subsequent* experience. (Having arrived too late for incorporation by the editing room into the "final print" of the first stimulus experience, it would appear in the next installment.) Libet's data suggest a tremendously variable editing window: "The conditioning cortical stimulus could be started more than 500 msec following the skin pulse and still modify the skin sensation, although in most cases retroactive effects were not observed with S-C intervals greater than 200 msec" (1981, p. 185). Libet is careful to define neuronal adequacy in terms of effects on subsequent unhurried verbal report: "The subject was asked to report, within a few seconds after the delivery of each pair of . . . stimuli" (1979, p. 195), and he insists that "the timing of a subjective experience must be distinguished from that of a behavioral response (such as in reaction time), which might be made before conscious awareness develops" (1979, p. 193).

This proviso permits him to defend a rival interpretation of Churchland's data. Churchland (1981a) attempted to discredit Libet's claim about the long rise time to

"neuronal adequacy" for consciousness by asking subjects in an experiment to say "go" as soon as they were conscious of such a skin stimulus as those used by Libet. She reported a mean response time over 9 subjects of 358 msec, which, she argued, showed that the subjects must have achieved neuronal adequacy by the 200 msec mark at the latest (allowing time for the production of a verbal response). Libet's reply is Stalinesque: A verbal reaction can be unconsciously initiated. "There is nothing magical or uniquely informative when the motor response is a vocalization of the word 'go' instead of the more usual one of a finger tapping a button. . . . The ability to detect a stimulus and to react to it purposefully, or be psychologically influenced by it, without any reportable conscious awareness of the stimulus, is widely accepted" (Libet 1981, pp. 187–88). And to the objection, "But what did Churchland's subjects think they were doing, if not saying, as requested, just when they were conscious of the stimulus?" Libet could give the standard Stalinesque reply: They did indeed eventually become conscious of the stimulus, but by then, their verbal report had already been initiated.[10]

For this reason Libet rejects such reaction time studies as Churchland's as having "an uncertain validity as a primary criterion of a subjective experience" (1981, p. 188). He favors letting the subject take his time: "The report is made unhurriedly within a few seconds after each trial, allowing the subject to introspectively examine his evidence" (p. 188). How, then, can he deal with the rival prospect that this leisurely pace gives the Orwellian revisionist in the brain plenty of time to replace the *veridical* memories of consciousness with *false* memories? "Reporting after the trial of course requires that processes of short-term memory and recallability be operative, but this presents no difficulty for subjects with no significant defects in these abilities" (Libet, p. 188).

This begs the question against the Orwellian, who is prepared to explain a variety of effects as the result of *normal* misremembering or hallucinatory recall, in which a prior, real event in consciousness is obliterated and replaced by subsequent memories. (For related discussions, see Allport 1988, pp. 171–76; Bisiach 1988, pp. 110–12.) Has Libet let the stew cook too long, or has Churchland sampled it too soon? If Libet wants to claim a *privileged* status for his choice of probe time, he must be prepared to combat the counterarguments.

Libet comes close to pleading *nolo contendere*: "Admittedly, a report of relative timing order cannot, in itself, provide an indicator of the 'absolute' time (clock-time) of the experience: As suggested, there is no known method to achieve such an indicator" (1981, p. 188). This echoes his earlier remark that there seemed to be "no method by which one could determine the absolute timing of a subjective experience" (Libet et al. 1979, p. 193). What Libet misses, however, is the possibility that this is because there is no such moment of absolute time (cf. Harnad, unpublished; 1989).

Churchland too fails to distinguish time represented from time of representing, in her criticisms (1981a; 1981b): "The two hypotheses differ essentially on just when the respective sensations *were felt* [our emphasis]," (1981a, p. 177) and

> Even if it be supposed that the sensations arising from the simultaneous skin and LM [medial lemniscus] sensations are *felt at exactly the same time* [our emphasis], the delay in neuronal adequacy for skin stimuli may well be an artifact of the setup. (1981b, p. 494)

Suppose that all such artifacts were eliminated, and *still* the sensations are "felt at exactly the same time." Will this mean that there is a time t such that stimulus 1 is felt at t and stimulus 2 is felt at t (the anti-materialist prospect) or only that stimulus 1 and stimulus 2 are felt as (experienced as) simultaneous? Churchland doesn't discourage the inference that Libet's findings, if vindicated, would wreak havoc (as he claims) on materialism. Elsewhere, however, she correctly notes that "intriguing as temporal illusions are, there is no reason to suppose there is something preternatural about them, and certainly there is nothing which distinguishes them from spatial illusions or motion illusions as uniquely bearing the benchmark of a nonphysical origin" (1981a, p. 178). This could only be the case if temporal illusions were phenomena in which *time was misrepresented*; if the *misrepresentings* take place at the "wrong" times, something more revolutionary is afoot.

Where does this leave Libet's experiments with cortical stimulation? As an interesting but inconclusive attempt to establish something about *how the brain represents temporal order*. Primary evoked potentials may somehow serve as specific reference-points for neural representations of time, although Libet has not shown this, as Churchland's technical criticisms make clear. Alternatively, the brain keeps its representations of time more labile. We don't represent seen objects as existing on the retina, but rather as various distances in the external world. Why should the brain not also represent events as happening *when* it makes the most "ecological" sense for them to happen? When we are engaged in some act of manual dexterity, "fingertip time" should be the standard; when we are conducting an orchestra, "ear time" might capture the registration. "Primary cortical time" might be the default standard (rather like Greenwich Mean Time for the British Empire) – a matter, however, for further research.

The issue has been obscured by the fact that both proponent and critic have failed to distinguish consistently between time of representing and time represented. They talk past each other, with Libet adopting a Stalinesque position and Churchland making the Orwellian countermoves, both apparently in agreement that there is a fact of the matter about exactly when (in "absolute" time as Libet would put it) a conscious experience happens.[11]

3.2. Libet's claims about the "subjective delay" of consciousness of intention. The concept of the absolute timing of an experience is exploited in Libet's later experiments with "conscious intentions," in which he seeks to determine their absolute timing experimentally by letting the subjects, who alone have direct access (somehow) to their experiences, do *self-timing*. He asked subjects to look at a clock (a spot of light circling on an oscilloscope) *while* they experience consciously intending, and to make a judgment about the position on the clock of the spot at the onset of intention, a judgment they can later, at their leisure, *report*.

Libet is clearer than most of his critics about the importance of keeping content and vehicle distinguished:

"One should not confuse *what* is reported by the subject with *when* he may become introspectively aware of what he is reporting" (Libet 1985a, p. 559). He recognizes (p. 560), moreover, that a judgment of simultaneity need not itself be simultaneously arrived at or rendered; it might mature over a long period of time (consider, for instance, the minutes it may take the stewards at the race track to develop and then examine the photo-finish picture on which they eventually base their judgment of the winner or a dead heat).

Libet gathered data on two time series: (1) the objective series, which includes the timing of the external clock and the salient neural events: the readiness potentials (RPs) and the electromyograms (EMGs), and (2) the subjective series (as later reported), which consists of mental imagery, memories of any preplanning, and, crucially, of a single benchmark datum for each trial: a simultaneity judgment of the form: *My conscious intention (W) began simultaneously with the clock spot in position P.*

Libet seems to have wanted to approximate the elusive *acte gratuit* discussed by the existentialists (e.g., Gide 1948; Sartre 1943), the purely motiveless – and hence in some special sense "free" – choice, and as several commentators have pointed out (Breitmeyer 1985; Bridgeman 1985; Danto 1985; Jung 1985; Latto 1985) such highly unusual actions (what might be called acts of deliberate pseudorandomness) are hardly paradigms of "normal voluntary acts" (Libet 1987, p. 784). But has he in any event isolated a variety of conscious experience, however characterized, that can be absolutely timed by such an experimental design?

He claims that when conscious intentions to act (at least of his special sort) are put into registration with the brain events that actually initiate the acts, there is an offset: Consciousness of intention lags 300–500 msec behind the relevant brain events. This does look ominous to anyone committed to the principle that "our conscious decisions" *control* our bodily motions. It looks as if *we* are located in Cartesian theaters where we are shown, with a half-second tape delay, the *real* decision-making that is going on *elsewhere* (somewhere *we* aren't). We are not quite "out of the loop" (as they say in the White House), but because our access to information is thus delayed, the most we can do is intervene with last-moment "vetoes" or "triggers." One who accepts this picture might put it this way: "Downstream from (unconscious) command headquarters, I take no real initiative, am never in on the birth of a project, but do exercise a modicum of executive modulation of the formulated policies streaming through my office."

This picture is compelling but incoherent. For one thing, such a "veto" would itself have to be a "conscious decision," it seems, and hence ought to require its own 300–500 msec cerebral preparation – unless one is assuming outright Cartesian dualism (see MacKay, 1985, who makes a related point). Setting that problem aside, Libet's model, as before, is Stalinesque, and the obvious Orwellian alternative is raised by Jasper (1985), who notes that both epileptic automatisms and behaviors occurring under the effect of such drugs as scopolamine show that "brain mechanisms underlying awareness may occur without those which make possible the recall of this awareness in memory afterward." Libet concedes that this "does present a problem, but was not experimentally testable" (p. 560).[12]

Given this concession, is the task of fixing the absolute *micro*timing of consciousness ill-conceived? Neither Libet nor his critics draw that conclusion. Libet, having carefully distinguished content from vehicle – *what* is represented from *when* it is represented – nonetheless tries to draw inferences from premises about what is represented to conclusions about the absolute timing of the representing in consciousness (cf. Salter 1989). Wasserman (1985) sees the problem: "The time when the external objective spot occupies a given clock position can be determined easily, but this is not the desired result." But he then falls into the Cartesian trap: "What is needed is the time of occurrence of the internal brain-mind representation of the spot."

"*The* time of occurrence" of the internal representation? Occurrence where? There is essentially continuous representation of the spot (representing it to be in various different positions) in various different parts of the brain, starting at the retina and moving up through the visual system. The brightness of the spot is represented in some places and times, its location in others, and its motion in still others. As the external spot moves, all these representations change, in an asynchronous and spatially distributed way. Where does "it all come together at an instant in consciousness"? Nowhere. Wasserman correctly points out that the task of determining where the spot was at some time in the subjective sequence is itself a voluntary task, and initiating it presumably takes some time. This is difficult not only because it is in competition with other concurrent projects (as stressed by Stamm 1985, p. 554), but also because it is unnatural – a conscious judgment of temporality of a sort that does not normally play a role in behavior control, and hence has no natural meaning in the sequence. The process of interpretation that eventually fixes the judgment of subjective simultaneity is itself an artifact of the experimental situation, and *changes the task*, therefore telling us nothing of interest about the actual timing of normal representational vehicles anywhere in the brain.

Stamm likens the situation to Heisenbergian uncertainty: "Self-monitoring of an internal process interferes with that process, so that its precise measurement is impossible" (p. 554). This observation betrays a commitment to the mistaken idea that *there is* an absolute time of intersection, "precise measurement" of which, alas, is impossible for Heisenbergian reasons (see also Harnad 1989). This could only make sense on the assumption that there is a particular privileged place where the intersection matters.

The all too natural vision that we must discard is the following: Somewhere deep in the brain an act-initiation begins; it starts out as an unconscious intention, and slowly makes it way to the theater, picking up clarity and power as it goes, and then, at an instant, *t*, it bursts on stage, where a parade of visual spot-representations are marching past, having made their way slowly from the retina, getting clothed with brightness and location as they moved. The audience or *I* is given the task of saying which spot-representation was "on stage" exactly when the conscious intention made its bow. Once identified, this spot's time of departure from the retina can be

calculated, as well as the distance to the theater and the transmission velocity. That way we can determine the exact moment at which the conscious intention occurred in the Cartesian Theater.

Some have thought that although that particular vision is incoherent, one does not need to give up the idea of absolute timing of experiences. There is an alternative family of models for the onset of consciousness that avoids the preposterousness of the Cartesian-centered brain. Couldn't consciousness be a matter not of arrival at a point but rather a matter of a representation exceeding some threshold of activation over the whole cortex or large parts thereof? On this model, an element of content becomes conscious at some time *t*, not by entering some functionally defined and anatomically located system, but by changing state right where it is: by acquiring some property or by having the intensity of one of its properties boosted above some criterial level.

The idea that content becomes conscious not by entering a subsystem, but by the brain's undergoing a state change of one sort or another has much to recommend it (see, e.g., Crick & Koch 1990; Kinsbourne 1988; Neumann 1990). Moreover the simultaneities and sequences of such mode-shifts can presumably be measured by outside observers, providing, in principle, a unique and determinate sequence of contents attaining the special mode. But this is still the Cartesian Theater if it is claimed that the real ("absolute") timing of such mode shifts is definitive of subjective sequence. The imagery is different, but the implications are the same. Conferring the special property that makes for consciousness at an instant is only half the problem; discriminating that the property has been conferred at that time is the other, and although scientific observers with their instruments may be able to do this with microsecond accuracy, how is the brain to do this? We human beings do make judgments about simultaneity and sequence among elements of our own experience, some of which we express, so at some point or points in our brains the corner must be turned from the actual timing of representations to the representation of timing. This is a process that takes effort in one way or another (Gallistel 1990), and wherever and whenever these discriminations are made, thereafter the temporal properties of the representations embodying those judgments are not constitutive of their content.

Suppose that a succession of widely spread activation states, with different contents, sweeps over the cortex. The actual, objectively measured simultaneities and sequences in this broad field are of no functional relevance *unless they can also be accurately detected by mechanisms in the brain.* What would make *this* sequence the stream of consciousness if the brain could not discern the sequence? What matters, once again, is not the temporal properties of the representings, but the temporal properties *represented*, something determined by how they are "taken" by subsequent processes in the brain.

3.3. Grey Walter's experiment: A better demonstration of the central contention of the Multiple Drafts model. It was noted above that Libet's experiment created an artificial and difficult judgmental task that robbed the results of the hoped-for significance. This can be brought out more clearly by comparing it to a similar experiment by Grey Walter (1963), with patients in whose motor cortex he had implanted electrodes. He wanted to test the hypothesis that certain burst of recorded activity were the initiators of intentional actions, so he arranged for each patient to look at slides from a carousel projector. The patient could advance the carousel at will, by pressing the button on the controller. (Note the similarity to Libet's experiment: This was a "free" decision, timed only by an endogenous rise in boredom, or curiosity about the next slide, or distraction, or whatever.) Unbeknownst to the patient, however, the controller button was a dummy, not attached to the slide projector at all. What actually advanced the slides was the amplified signal from the electrode implanted in the patient's motor cortex.

One might suppose that the patients would notice nothing out of the ordinary, but in fact they were startled by the effect, because it seemed to them as if the slide projector was anticipating their decisions. They reported that just as they were "about to" push the button, but before they had actually decided to do so, the projector would advance the slide – and they would find themselves pressing the button with the worry that it was going to advance the slide twice! The effect was strong, according to Grey Walter's account, but apparently he never performed the dictated followup experiment: introducing a variable delay element to see how large a delay had to be incorporated into the triggering to eliminate the "precognitive carousel" effect.

An important difference between Grey Walter's and Libet's designs is that the judgment of temporal order that leads to surprise in Grey Walter's experiment is part of a normal task of behavior monitoring. In this regard it is like the temporal order judgments by which our brains distinguish moving left-to-right from moving right-to-left, rather than "deliberate, conscious" order judgments. The brain in this case has set itself to "expect" visual feedback on the successful execution of its project of advancing the carousel, and the feedback arrives earlier than expected, triggering an alarm. This could show us something important about the actual timing of content vehicles and their attendant processes in the brain, but it would not, contrary to first appearances, show us something about the "absolute timing of the conscious decision to change the slide."

Suppose, for instance, that an extension of Grey Walter's experiment showed that a delay as long as 300 msec (as implied by Libet) had to be incorporated into the implementation of the act in order to eliminate the subjective sense of precognitive slide-switching. What such a delay would in fact show would be that expectations set up by a decision to change the slide are tuned to expect visual feedback 300 msec later, and to report back with alarm under other conditions. The fact that the alarm eventually gets interpreted in the subjective sequence as a perception of misordered events (change before button push) shows nothing about *when* in real time the consciousness of the decision to press the button first occurred. The sense the subjects reported of not quite having had time to "veto" the initiated button push when they "saw the slide was already changing" is a natural interpretation for the brain to settle on (eventually) of the various contents made available at various times for incorporation into the narrative. Was this sense already there

at the first moment of consciousness of intention (in which case the effect requires a long delay to "show time" and is Stalinesque) or was it a retrospective reinterpretation of an otherwise confusing *fait accompli* (in which case it is Orwellian)? This question should no longer seem to demand an answer.

4. Conclusion

The Multiple Drafts model has many other implications for scientific theories of consciousness (Dennett 1991b), but our main conclusion in this target article is restricted to temporal properties of experience: The representation of sequence in the stream of consciousness is a product of the brain's interpretative processes, not a direct reflection of the sequence of events making up those processes. Indeed, as Jackendoff has pointed out to us, what we are arguing for in this essay is a straightforward extension to the experience of time of the common wisdom about the experience of space; the representation of space in the brain does not always use space-in-the-brain to represent space, and the representation of time in the brain does not always use time-in-the-brain. It may be objected that the arguments presented here are powerless to overturn the still obvious truth that our experiences of events occur in the very same order that we experience them to occur. If someone thinks the thought, "One, two, three, four, five," his thinking "one" occurs before his thinking "two" and so forth. The example does illustrate a thesis that is true in general and does indeed seem unexceptioned, so long as we restrict our attention to psychological phenomena of "ordinary," macroscopic duration. But the experiments we selected for discussion are concerned with events that were constricted by unusually narrow timeframes of a few hundred milliseconds. At this scale, we have argued, the standard presumption breaks down.

It might be supposed, then, that we are dealing only with special cases. These limiting cases may interestingly reveal how the brain deals with informational overload, but, one might suggest, they are unrepresentative of the brain's more usual manner of functioning. The contrary is the case, however, as might be anticipated, in view of the brain's well-known propensity for applying a limited number of basic mechanisms across a wide range of situations. The processes of editorial revision that are dramatically revealed in the time-pressured cases continue indefinitely as the brain responds to the continued demands of cognition and control. For instance, as time passes after an event has occurred, that event may be recalled to episodic memory, but to an ever more limited extent. After some days, an occurrence that may have unrolled over minutes or more is remembered within as restricted a time frame as those we have been discussing. Such memories present not as randomly blurry or depleted versions but as internally coherent, simplified renderings of what are taken to be the most important elements. Temporal succession is typically an early victim of this reorganization of the event, sacrificed in favor of (apparently) more useful information (as instanced in the phi phenomenon).

We perceive – and remember – perceptual events, not a successively analyzed trickle of perceptual elements or attributes locked into succession as if pinned into place on a continuous film. Different attributes of events are indeed extracted by different neural facilities at different rates, (e.g., location vs. shape vs. color) and people, if asked to respond to the presence of each one in isolation, would do so with different latencies, depending on which it was, and on other well-explored factors. The relative timing of inputs plays a necessary role in determining the information or content of experience, but it is not obligatorily tied to any stage or point of time during central processing. How soon we can respond to one in isolation, and how soon to the other, does not exactly indicate what will be the temporal relationship of the two in percepts that incorporate them both.

There is nothing theoretically amiss with the goal of acquiring precise timing information on the mental operations or informational transactions in the brain (Wasserman & Kong 1979). It is indeed crucial to developing a good theory of the brain's control functions to learn exactly when and where various informational streams converge, when "inferences" and "matches" and "bindings" occur. But these temporal and spatial details do not tell us directly about the contents of consciousness. The temporal sequence *in consciousness* is, within the limits of whatever temporal control window bounds our investigation, purely a matter of the content represented, not the timing of the representing.

ACKNOWLEDGMENTS

The original draft of this essay was written while the authors were supported by the Rockefeller Foundation as Scholars in Residence at the Bellagio Study Center, Villa Serbelloni, Bellagio, Italy, April, 1990. We are grateful to Kathleen Akins, Peter Bieri, Edoardo Bisiach, William Calvin, Patricia Churchland, Robert Efron, Stevan Harnad, Douglas Hofstadter, Tony Marcel, Odmar Neumann, Jay Rosenberg, and David Rosenthal for comments on subsequent drafts.

NOTES

1. A philosophical exception is Vendler (1972; 1984) who attempts to salvage Cartesian dualism. A scientific exception is Eccles (e.g., Popper & Eccles 1977).

2. What about the prospect of a solitary Robinson Crusoe scientist who performs all these experiments wordlessly on himself? Would the anomalies be apparent to this lone observer? What about reconstructing these experiments with languageless animals? Would we be inclined to interpret the results in the same way? Would we be justified? These are good questions, but their answers are complicated, and we must reserve them for another occasion.

3. Such a "postmark" can be in principle be added to a vehicle of content at any stage of its journey; if all materials arriving at a particular location come from the same place, by the same route at the same speed, their "departure time" from the original destination can be retroactively stamped on them, by simply subtracting a constant from their arrival time at the way station. This is an engineering possibility that is probably used by the brain for making certain automatic adjustments for standard travel times.

4. "The essence of much of the research that has been carried out in the field of sensory coding can be distilled into a single, especially important idea – any candidate code can represent any perceptual dimension; there is no need for an isomorphic relation between the neural and psychophysical data. Space can represent time, time can represent space, place can represent quality, and certainly, nonlinear neural functions can represent linear or nonlinear psychophysical functions equally well" (Uttal 1979). This is a widely acknowledged idea,

but, as we will show, some theorists (mis-)understand it by tacitly reintroducing the unnecessary "isomorphism" in a dimly imagined subsequent translation or "projection" in consciousness.

5. Cf. Pylyshyn 1979: "No one . . . is disposed to speak *literally* of such physical properties of a mental event as its color, size, mass, and so on – though we *do* speak of them as *representing* (or having the experiential content of) such properties. For instance, no one would not properly say of a thought (or image) that it was large or red, but only that it was a thought *about* something large or red (or that it was an image *of* something large or red). . . . It ought to strike one as curious, therefore, that we speak so freely of the *duration* of a mental event."

6. P. S. Churchland (1981a, p. 172) notes a difference between "masking in the usual sense" and "blanking in short term memory," which perhaps is an allusion to these two possibilities, but does not consider how one might distinguish between them.

7. Consider the medio-temporal region of cortex (MT), which responds to motion (and apparent motion). Suppose then that some activity in MT *is* the brain's concluding that there was intervening motion. There is no further question, on the Multiple Drafts model, of whether this is a pre-experiential or post-experiential conclusion. It would be a mistake to ask, in other words, whether this activity in MT was a "reaction to a conscious experience" (by the Orwellian historian) as opposed to a "decision to represent motion" (by the Stalinesque editor).

8. See also his dismissal of MacKay's suggestion of a more moderate reading (Libet 1981, p. 195; 1985b, p. 568).

9. Libet's final summation in 1981, on the other hand, was inconclusive: "My own view . . . has been that the temporal discrepancy creates relative difficulties for identity theory, but that these are not insurmountable" (p. 196). Presumably they would be undeniably insurmountable on the backwards *projection* interpretation, and Libet later (1985b, p. 569) describes these difficulties in a way that seems to require the milder reading: "Although the delay-and-antedating hypothesis does not separate the actual time of the experience from its time of neuronal production, it does eliminate the necessity for simultaneity between the *subjective timing* of the experience and the actual clock-time of the experience." Perhaps Eccles's enthusiastic support for a radical, dualistic interpretation of the findings has misdirected the attention of Libet (and his critics) from the mild thesis he sometimes defends.

10. In an earlier paper, Libet conceded the possibility of Orwellian processes and supposed there might be a significant difference between unconscious mental events and conscious-but-ephemeral mental events: "There may well be an immediate but ephemeral kind of experience of awareness which is not retained for recall at conscious levels of experience. If such experiences exist, however, their content would have direct significance only in later unconscious mental processes, although, like other unconscious experiences, they might play an indirect role in later conscious ones" (1965, p. 78).

11. Harnad (1989) sees an insoluble problem of measurement, but denies our contention that there is no fact of the matter: "Introspection can only tell us when an event *seemed* to occur, or which of two events *seemed* to occur first. There is no independent way of confirming that the real timing was indeed as it seemed. Incommensurability is a methodological problem, not a metaphysical one." So Harnad asserts what we deny: that among the real timings of events in the brain is a "real timing" of events *in consciousness*.

12. In a later response to a similar suggestion of Hoffman and Kravitz (1987) Libet asks the rhetorical question, "Are we to accept the primary evidence of the subjects' introspective report (as I do), or are we going to insist that the subject had a conscious experience which he himself does not report and would even deny having had?" (1987, p. 784). This is another expression of Libet's a priori preference for a Stalinesque position.

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

The where and when of what?

Michael V. Antony

Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139
Electronic mail: *mvantony@athena.mit.edu*

Dennett & Kinsbourne (D & K) set out to replace what they take to be a bad picture of consciousness, "Cartesian materialism," with one they prefer, their "Multiple Drafts" model. Once made explicit, they maintain, Cartesian materialism is an obvious mistake (sect. 1.1., para. 10). Nevertheless it is hard to escape its grip, they claim; its imagery continues to infect much current thought about consciousness. In this commentary I shall not discuss D & K's arguments against Cartesian materialism in any detail. Instead, I shall argue that the most charitable reading of their critique commits them to the view that consciousness *does not exist*–to *eliminativism* about consciousness. Although eliminativism follows from their central claims, other passages suggest that they believe consciousness *does* exist. Consequently, their position on the ontological status of consciousness needs clarification.

Cartesian materialism is the view that remains when Cartesian dualism is rejected but "the associated imagery of a central (but material) theater where 'it all comes together'" is retained (sect. 1.1, para. 8). Although Cartesian materialism includes the questionable view that there is a single place in the brain where all conscious experiences occur, it also includes the more plausible and widely held view that conscious states, processes, and so forth, are *functionally characterizable* (sect. 1.1, para. 10; sect. 2.2, para. 1). According to functionalism, token brain states or processes are conscious if and only if they bear appropriate causal relations to other token states and processes (and inputs and outputs). Consequently, on functionalist theories there need be no single place in the brain that subserves only consciousness, let alone a single place where all consciousness occurs. Each experience must simply occur somewhere or other.

In arguing against Cartesian materialism, D & K set out to show that there are no facts about exactly when token experiences happen (note 12), for example, when they begin or end. Settling "on some moment of processing in the brain as the moment of consciousness," they write, "has to be arbitrary" (sect. 2.2, para. 21). There is no moment of absolute time of an experience (sect. 3.1, para. 8; sect. 3.2, para. 8). From the claims expressed in these passages, and others like them, it is meant to follow that Cartesian materialism is false. The idea, presumably, is that if there is a place in the brain where a given experience occurs, there is necessarily some time when it occurs also. But there is no such time and consequently no such place.

Putting aside the question of whether D & K have argued soundly for the claims expressed in the passages cited above, it is worth considering how those claims are to be best understood. I can think of three ways of interpreting their statements. The first takes them as straightforwardly denying that conscious experiences are temporally located, that they begin, persist for some interval, and end. Clearly this interpretation leads directly to eliminativism, given materialistic assumptions. For

every physical state, event, process, and so on, is temporally located in this sense. If there exist conscious states and processes, therefore, they are as well.

Perhaps, however, D & K's picture is a more subtle one. What they seem to be critical of is the notion that experiences have "exact" or "absolute" temporal locations, that they have beginnings or endings that can be fixed upon with "precision." Perhaps, then, their idea is that while conscious experiences are temporally located in the sense in which everything is, the temporal boundaries of conscious experiences are *vague*. Thus, if one wished to state precisely (e.g., in milliseconds) when an experience began or ended, one would have to choose arbitrarily from among equally good alternatives, there being no fact of the matter about which choice is correct. The trouble with this interpretation, however, is that it is trivially true, since *everything* has vague temporal boundaries relative to some sufficiently fine-grained scale or other. Consider, for example, a scale of microseconds (one millionth of a second), and everyday events like a sneeze, the running of an engine, or an action potential in a cortical cell. Fixing the beginnings or endings of any of these events in microseconds would demand arbitrary choices from among thousands of alternatives. The temporal boundaries not only of experiences, therefore, but of every neurophysiological state or event are vague relative to some temporal scale. (Analogous points hold for spatial boundaries.) Consequently, merely asserting that the temporal boundaries of experiences are vague amounts to a truism, and has no bearing on Cartesian materialism.

A third way of reading the above passages is to understand them as claiming that conscious experiences are temporally located, but the time scale appropriate for the measurement of neurophysiological events (e.g., milliseconds) is too fine-grained to determine nonarbitrary temporal boundaries for experiences. This claim is not trivial since it implies that conscious experiences cannot be identified with functional states or processes (assuming such states or processes are realized by neurophysiological states or processes). The basic idea underlying this interpretation, then, is that the temporal boundaries of conscious experiences are "bigger and fuzzier" than the boundaries that carve up functional states and processes in the brain. But now there is a difficulty: Conscious experiences have no role in the functional organization of the brain. Hence they are not causally related to any states or processes that *are* functionally characterizable. (If they were they would be functional states or processes themselves.) It follows that conscious experiences cannot be introspected and reported on, they cannot be caused by sensory input, they cannot play a role in causing behavior and thought, and so on. Indeed it is difficult to see how one could even *think* about one's own conscious experiences, if one's thoughts were causally unconnected to those experiences. It would seem, therefore, that this third interpretation reduces to absurdity.

It appears that the only way D & K's remarks on the temporal properties of experiences an be understood as expressing an intelligible and substantive claim is by interpreting them in such a way that eliminativism about consciousness is entailed. The two alternative interpretations that seem possible are either trivial or incoherent. The trouble with taking D & K as advocating eliminativism, however, is that nowhere do they explicitly endorse eliminativism. On the contrary, their target article contains numerous passages that imply that they think consciousness does exist; for instance, their last sentence: "The temporal sequence *in consciousness* is . . .purely a matter of the content represented, not the timing of the representing." (conclusion, para. 4). Thus if Dennett & Kinsbourne believe consciousness exists, they must provide an interpretation of their remarks on the temporal properties of experience that is distinct from the three I have discussed. If they do not, the passages suggesting they do must be clarified.

Throwing the conscious baby out with the Cartesian bath water

J. Aronson, E. Dietrich, and E. Way

Program in Philosophy, Computer, and Systems Science, Department of Philosophy, State University of New York Binghamton, Binghamton, NY 13905

Electronic mail: *dietrich@bingvaxu.cc.binghamton.edu*

Ms. Molly Bloom drifts off to sleep, consciously thinking these last thoughts ". . . all perfume yes and his heart was going like mad and yes I said yes I will Yes." When Ms. Bloom trades consciousness for sleep, she is trading away a single, unified narrative for none. This is what Joyce believed, this is what we believe, but this isn't what Dennett & Kinsbourne (D & K) believe. Instead, D & K believe that there is no unified stream of consciousness, no place (even functionally speaking, see sect. 1.1, para. 8) where "it all comes together." They want to replace the notion of consciousness as a master discriminator with braided streams of consciousness (to use a geological metaphor, the only one that hasn't been used yet). Instead of a single narrative, they want a "a skein of narratives."

D & K have developed their view of consciousness, called the "multiple drafts model," in order to avoid Cartesian materialism and its associated view of consciousness: the Cartesian Theater model. There are several things we like about this target article. We applaud D & K's vigilance against Cartesianism (this pernicious doctrine has more lives than a cat; we tolerate even its mild forms to our detriment); we agree that there is no one place in the brain where consciousness occurs, and we do think that their clever Orwellian-Stalinesque argument shows that there is an unsolvable *descriptive* problem in the study of consciousness. Furthermore, the multiple processing and time-stamping aspects of their multiple drafts model are important additions to cognitive science's growing understanding of consciousness. But we think that they have thrown the baby out with the bath water: They seem to have rejected the essential phenomenology of consciousness in an attempt to avoid dualism and implausible neurophysiology. At least, this shall be our plaint.

D & K reject two different versions of the Cartesian Theater model: an anatomical version and a functional version. The anatomical version is the materialistic cousin of Descartes' original dualism. Descartes thought the pineal gland was the gateway to the conscious mind. Rejecting Descartes' dualism but keeping the anatomical sentiments, one might hypothesize that the pineal gland is the seat of consciousness; it is where consciousness takes place. But, as we learn from the neuroanatomists, neither in the pineal gland nor probably anywhere else in the brain is the seat of consciousness. Consciousness, as D & K note, is distributed. However, they think that rejecting the anatomical version of the Cartesian Theater requires rejecting the functional version as well (sect. 1.1, para. 8). This is a mistake. There could well be a functionally defined locus of consciousness that is both physically distributed and operates at different levels and times. One can believe as they do that the subjective temporal properties of events are *not* determined by the objective temporal properties of events anywhere in the brain and yet still believe that "it all comes together somewhere – functionally speaking." Put another way, one can believe that contents are temporally stamped – that they carry their temporal representations around with them – without having to give up the notion that consciousness is a single unified narrative.

There is more to this point than mere logic. Phenomenologically, it certainly seems that consciousness is unified and that the stream of consciousness is a single narrative (it's not always coherent, but it is a single narrative: Ms. Bloom's nocturnal thoughts are the result of *subconscious* competing and parallel processes). We will need much more than what we've been

given by D & K before we surrender this perception. For starters, they owe us an explanation of why we have this perception in the first place.

Our criticism of D & K as well as our positive position is summed up in Figure 1. Figure 1a represents the position we take D & K to be attacking, and we agree that it is incorrect. Figure 1b represents our understanding of the multiple drafts model. As one can tell from Figure 1b, the multiple drafts model seems to predict that each person has multiple conscious selves, each independent from one another. This seems wrong: There is no logical reason we should adopt such a view, nor do the experiences of individuals support it. Furthermore, such a view could be construed as the very one D & K are attacking, namely the Cartesian Theater – instead of just one theater, there are several . . . a sort of theater district.

The view which captures what we like about D & K's ideas while remaining faithful to the phenomenology is Figure 1c (the functional theater). This model does not require a single location for the seat of consciousness but still allows for a single stream of consciousness, one which is constantly revised and updated, but is nonetheless unified.

Finally, the functional theater model is not open to the charge of dualism because it could be implemented as a computer program. For example, one kind of metaprocessor is a machine that takes as input the run-time trace of an object-level processor and outputs information implicit in that run-time trace. One kind of information that is implicit in this sense is the data type of the object-level process. (Determining that a trace is a trace of a stack growing and shrinking is an example of the kind of thing

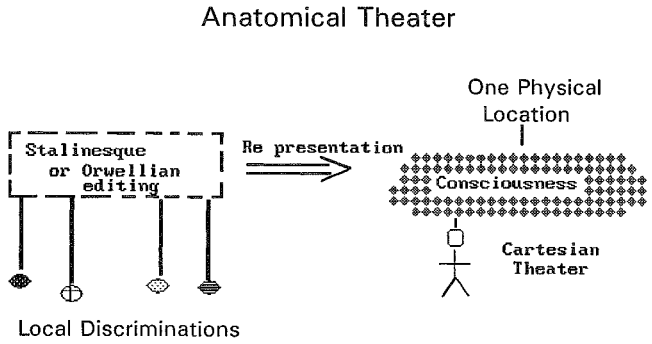

Figure 1a (Aronson et al.). The view that D & K are correctly attacking, in which consciousness *is* events derived from local discriminations displayed at a specific point in the brain, subtly implies the existence of a homunculus.

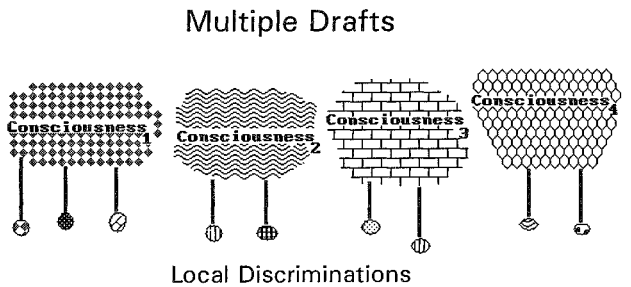

Figure 1b (Aronson et al.). D & K's model of consciousness: the Multiple Drafts model. This view implies that there are multiple conscious selves in each person, a view that seems intuitively implausible.

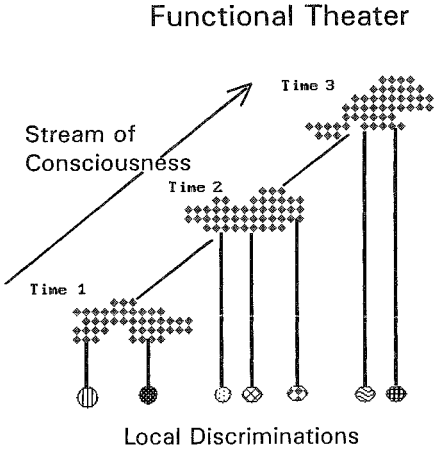

Figure 1c (Aronson et al.). Like D & K's Multiple Drafts model, the functional theater model does not require consciousness to be located at a specific place, but still allows for consciousness to be unified.

such a metaprocessor could do.) If we hypothesize that thinking is the execution of various (instantiated) data types (this hypothesis is one way of being a computationalist), then we can describe consciousness as a metaprocessor that makes explicit the implicit data type a current thought is, for example, the *belief that snow is cold* or the *desire for ice cream* (see Dietrich 1985).

We conclude that while D & K are clear on what the multiple drafts model is *not* (namely, the anatomical version of the Cartesian Theater), they are not at all clear on what their model is. And, they are stuck with a dilemma. If they are trying to explain consciousness as it *appears*, and they therefore do hold with some kind of unifying process, then they still owe us an explanation of how the parallel and conflicting streams come together; and talk of time stamping isn't a sufficient explanation. On the other hand, if they deny that consciousness is unified, then it seems that they have not explained consciousness at all, but rather simply thrown it out.

Consciousness is associated with central *as well as* distributed processes

Bernard J. Baars[a] and Michael Fehling[b]

[a]The Wright Institute, Berkeley, CA 94704[1] and [b]Department of Engineering-Economic Systems, Stanford University, Stanford, CA 94305-4025

Electronic mail: *[a]baars@bayes.stanford.edu*

Dennett & Kinsbourne (D & K) present a strong case that a naive "point center" conception of consciousness is wrong. However, their Multiple Drafts model seems to deny for consciousness any integrative or executive function at all. Yet if consciousness has no "central" function, why would the authors want to bring their ideas to the consciousness of this audience? They would surely be unhappy if all potential readers were unable to pay conscious attention to their work, for example, by listening to the radio while trying to read the target article. This commonsense *reductio ad absurdum* makes a point that is also backed by much experimental evidence: Consciousness plays a central role in integrative and executive functions (see Baars,

1988, Chapter 2, and below). This central role, however, does not rule out the existence of multiple distributed systems.

An extreme dichotomy between centrist and distributed views of consciousness, as if only these alternatives existed, is inimical to clear thought. These are not the only options. A more comprehensive approach can integrate both central and distributed functions in a single, coherent framework. One such framework has been worked out in detail by Baars (1983; 1988). Fehling's Schemer II architecture (Fehling et al. 1990) supports a principled implementation of this type of system with great ease and flexibility.

Consider the extensive evidence about comparable conscious and unconscious processes shown in Table 1, a highly condensed summary of the extensive experimental literatures on immediate memory, selective attention, automaticity, errors, and directly observed brain processes. Consciousness appears to be closely associated with a limited capacity system, as is well known. But Table 1 also suggests that the limited-capacity system interacts constantly with a largely unconscious, highly distributed "society" of special-purpose processes. At any single time, most of these distributed processes are not available to consciousness, nor are they under conscious control; yet they certainly help shape conscious experiences and are indispensible to the execution of conscious goals.

Table 1 suggests the existence of a *parallel-interactive* system, one in which a vast society of anatomical and functional systems in the brain is largely unconscious. Out of this great collection of specialized functions – which certainly involve multiple drafts of any particular input, as suggested by D & K – there emerges a serial flow of remarkably restricted size. It can hardly be accidental that this emergent serial flow happens to be intimately associated with conscious experience. Our work with the global-workspace architecture suggests that a serial component of a massive parallel system becomes necessary when many parallel distributed processors must *interact* in some maximally integrative fashion. Interactivity enforces seriality.

Without defending Cartesian Materialism, we want to emphasize several central facts that remain unexplained by D & K's Multiple Drafts model:

Table 1 (Baars & Fehling). *Contrastive features of comparable conscious and unconscious mental processes*

Conscious processes	Unconscious processes
1. Computationally inefficient. High error rate, relatively low speed, and mutual interference between conscious processes.	1. Very efficient in routine tasks. Low error rate, high speed, and little mutual interference.
2. Great range of contents. Great ability to relate different conscious contents to each other. Great ability to relate conscious events to their unconscious contexts.	2. Each routine process has a limited range of contents. Each routine process is relatively isolated and autonomous. Each routine process is relatively context-free.
3. High internal consistency at any single moment, seriality over time, and limited processing capacity.	3. The set of routine, unconscious processes is diverse, can sometimes operate in parallel, and together have great processing capacity.

1. *Conscious experience is internally consistent.* For example, potentially ambiguous percepts are always experienced unambiguously at any particular moment, although there is much reason to suppose that unconscious representations of the ambiguous stimulus co-exist with the internally consistent conscious one.

2. *There is a need for some degree of functionally central control and coordination of multiple, parallel lines of deliberation and action.*

3. *Some particular stream of processing very quickly comes to predominate in any particular context.*

4. *The dominant processing stream often incorporates and refines information from other parallel streams.*

5. *An agent's ability to select among multiple streams is limited by certain integrative constraints, such as the need to act in a non-self-contradictory way.* In driving a car, we do not step on the accelerator and the brake pedal at the same time. Our actions are somehow integrated in the service of global objectives, although the details of those actions may be relegated to distributed systems.

Our current thinking embodies the principal mechanisms specified by Baars' Global Workspace (GW) theory (1988), which suggests that conscious experience reflects the workings of a global "broadcasting system," analogous to a bulletin board or blackboard. The contents of a limited-capacity GW in our architecture are sufficient to initiate, coordinate, and integrate multiple, otherwise independent processing streams. GW data must include descriptions of properties of the processes themselves as well as representations of data returned by those processes. No confusion need arise in this architecture between the onset time of a process constructing a temporal representation and the time value encoded as a result. Both data produced by processes and data describing process behavior can be encapsulated within process streams, or within the globally accessible GW, or both, as required. Local coordination among process streams can be achieved by direct process-to-process communication as well as indirectly via the contents of the GW. Consequently, this architecture manifests hierarchical functional organization that changes in response to changing needs for interaction among its constituent processes, as well as at the global level.

We are currently investigating such issues by means of a principled, computer-based functional architecture developed by Fehling et al. (1989). In this framework, consciousness, reflection, self-concept and self-awareness, as well as a host of unconscious processes, emerge as highly functional (e.g., Fehling et al. 1990). An agent's problem-solving, decision-making, and control of actions is accomplished by a large collection of distributed cognitive-process elements that operate independently and concurrently except where they must be coordinated. As most computer scientists who design distributed systems will attest, a significant degree of explicit global coordination of the distributed processes is needed for the cognitive system as a whole to act in a coherent and integrated manner. The Schemer architecture allows us to examine alternative ways to coordinate the actions of these highly distributed cognitive processes as well as the mechanisms that serve the need for integration.

In sum, though we have little disagreement with D & K's critique of Cartesian Materialism, it seems that the Multiple Drafts model errs on the side of an unnecessary extreme. It does not address the need for overall coordination of distributed processes and the manifestation of such coordination in conscious experience. Our own work suggests a less extreme alternative. Although a global workspace architecture does indeed use a functionally centralized "stream of consciousness," it appears quite able to handle the facts of temporal representation addressed by Dennett & Kinsbourne.

NOTE

1. For correspondence, please contact: Bernard J. Baars, The Wright Institute, 2728 Durant Ave., Berkeley, CA 94704; telephone 415-841-9230.

Begging the question against phenomenal consciousness

Ned Block

Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139
Electronic mail: *block@psyche.mit.edu*

Dennett & Kinsbourne's (D & K's) argument hinges on the unmentioned and unargued assumption that there is no such thing as phenomenal consciousness. Those readers who believe in phenomenal consciousness should not find the argument convincing.

What is phenomenal consciousness? There is a great chasm between those who think about the mind: On one side are those who accept a concept of consciousness distinct from any cognitive or information processing or functional notion; on the other side are those who reject any such concept. Dennett (1991b) gives a book-length argument against such a conception, drawing on, among other things, the raw materials of this target article, but there is no such argument in the target article.

The concept of consciousness at issue attracts descriptions like "raw feel," "immediate phenomenological quality," and Nagel's (1974) "what it is like." This is the concept of consciousness that leads us to speak of an "explanatory gap": At this stage in the relevant sciences we have no idea how the neural substrate of my pain can explain why my pain feels like *this* rather than some other way or no way at all. This is the concept of consciousness that gives rise to the famous "inverted spectrum" hypothesis – things we both call "green" look to you the way things we both call "red" look to me – and the "absent qualia qualia" hypothesis, the idea that there could be a machine that was computationally like us, but was nonetheless a phenomenally unconscious zombie. Note that these conundra are *routes* to phenomenal consciousness – they do not constitute it. One can accept phenomenal consciousness without accepting any of them because our fundamental access to phenomenal consciousness derives from our acquaintance with it.[1]

Of course, those who accept phenomenal consciousness do not disparage other concepts of consciousness. We can speak of information as conscious in the sense that it is inferentially promiscuous (Stich 1978), that is, it is easily available to be used as a premise in reasoning and in formulating plans, and it is available for reporting. (For example, information that is repressed is not inferentially promiscuous.) Or we can speak of a state as conscious in the sense that it is accompanied by a thought to the effect that one is having that state (Armstrong 1968; Rosenthal 1986). Or we can speak of a state as conscious in the sense that one can monitor it or have internal soliloquies about it.

The question begged. The point of the "multiple drafts" metaphor is that with the various forms of electronic quasi-publishing now flourishing, we may one day (though clearly not yet) have a situation in which any decision as to which of the many versions is to be counted as "the publication" will be arbitrary. D & K write that since a perception of an event is spread over the brain in space and therefore in time, labeling any of the stages or revisions conscious would be similarly arbitrary. *But what if some of the brain representations of an event making up this spatio-temporal volume are phenomenally conscious whereas others are not?* It is surely *non*arbitrary to label those phenomenally conscious events as conscious. D & K's claim that it is arbitrary to select some representations as conscious is plausible only if one swallows their unmentioned irrealism about phenomenal consciousness. Only if there is *no such thing* as phenomenal consciousness would it be arbitrary to ascribe "it" to one representation rather than another.

If there is such a thing as phenomenal consciousness, then presumably this fact will be reflected at the neural level. Perhaps there will be differences between those brain representations that are phenomenally conscious and those that are not of the sort that Crick and Koch (1990), have proposed. The D & K argument depends on supposing that the Crick & Koch research project is a confusion, a search for something that does not exist, but we are given no argument to this effect.

Note that the D & K arguments against Cartesian materialism (even assuming that they are successful) cast no doubt on the reality of phenomenal consciousness. The hypothesis I just mentioned, to the effect that some perceptual representations of an event are conscious whereas others are not, makes no commitment to there being any single place where all phenomenally conscious events live. Perhaps there are phenomenally conscious events in many different areas of the brain. No one endorses Cartesian materialism, but many of us endorse phenomenal consciousness. Note also that one who accepts phenomenal consciousness can also accept all sorts of borderline cases between consciousness and nonconsciousness, and other sorts of indeterminacies of consciousness; perhaps some of the stages of perceptual processing will be like phenomenally conscious events in some ways but not in others. Antirealists often try to stick realists with one or another version of the law of the excluded middle: "If there is such a thing as phenomenal consciousness, then every question about it has a definite yes or no answer." Is phenomenal event A before phenomenal event B? If one event contains the other, there will be no yes-or-no answer.

Orwellian versus Stalinist accounts. D & K repeatedly write that "nothing discernible to 'inside' or 'outside' observers could distinguish" Orwellian from Stalinist options (sect. 2.2, para. 29). Both claims are false if phenomenal consciousness exists, as I will now argue.

"Inside." On the Orwellian story about the lady who runs by, for example, there is a phenomenal experience of a lady with long hair and no glasses. On the Stalinist story, there is no such phenomenal experience. The difference between *some* such phenomenal experience and *no* such phenomenal experience is an "inside" difference if there ever was one. What D & K appear to mean by something discernible "inside" is something discernible in the subject's *judgments*. But to assume that the subjective is exhausted by judgments is to beg the question against phenomenal consciousness, which, if it exists, is *not* just a matter of judgments.

The question-begging nature of the argument is summed up by the following lines about the Orwellian/Stalinist disagreement: " . . . they both account for the subjective data – whatever is obtainable 'from the first person perspective' – because they agree about how it ought to 'feel' to subjects." (sect. 2.2, para. 19) On the contrary, the Orwellian and Stalinist theories do *not* agree about how things feel to the subject; the Orwellian accepts the phenomenally conscious experience of a lady with long hair and no glasses; the Stalinist rejects it. What the Orwellian and the Stalinist agree on is the subject's *judgments, not on what he feels.*

D & K may wish to retreat to the position that one cannot use introspection as a source of data to decide between Orwellian and Stalinist theories of one's own experiences. This is not something that believers in phenomenal consciousness should dispute, since we believe that our privileged access, such as it is, extends to our experiences themselves, but does not make our memories of them infallible. If we want to decide between Orwellian and Stalinist accounts, we will need to appeal to science, and to this matter we now turn.

"Outside." The D & K claim that "outside" observers could never tell the difference between Orwellian and Stalinist hypotheses is equally question-begging. If the presence or absence of phenomenal consciousness is a real property, ingenious experimenters will presumably find a way to get an experimental handle on it. (See Potter 1975 and 1976, for the beginnings of a line of evidence for an Orwellian stance.) D & K say that the Orwellian and the Stalinist can agree on where the mistaken content enters the causal pathway, they just disagree about

whether the content prior to the point at which the mistaken content is introduced is phenomenally conscious or not. They conclude that "this is a difference that makes no difference" (sect. 2.2, para. 19). But it makes no difference only on the assumption that there are no real facts about phenomenal consciousness, an assumption which the target paper does not state or support.

One crude approach to telling the difference between at least some Orwellian and Stalinist pairs of hypotheses is this: In some cases, the Orwellian postulates a change in memory representation not accepted by the Stalinist. (See for example, the case of Chase and Sanborn in Dennett 1988.) In such cases, if the relevant memory system could be independently isolated and found not to change, that would support the Stalinist. I tried this out on Dennett (he mentioned that it was independently suggested by P. S. Churchland); his initial response was that we could not rely on any independent identification of the relevant memory system. If D & K want to pursue this line of thought, they owe us a reason to believe that standard scientific procedures cannot succeed in making such an independent identification. The hypothesis that the earth is flat can be insulated from data by the ad hoc postulation of special forces that make a disk look like a sphere, for example. Why should we believe that standard scientific criteria cannot possibly distinguish between Orwellian and Stalinist hypotheses?

Another line of empirical investigation would be to isolate the neurological nature of phenomenal consciousness itself as in the Crick and Koch project mentioned earlier. Again, D & K owe us a reason to think that this cannot be done.

Mellor, Churchland, and Harnad. D & K correctly point out that the temporal order of outside events needn't be represented by the temporal order of inside events. This Kantian point (Kant distinguished apprehension of succession from succession of apprehension) is certainly correct. D & K accuse P. S. Churchland, and somewhat less directly, S. Harnad and H. Mellor, of confusing order of representeds with order of representings. In my view, the quotations that they give from Churchland (1981a; 1981b) and Harnad (1982) show no signs of such a confusion, but rather just a commitment to the reality of phenomenal consciousness. I suggest you look at the quotations (sect. 3.1, para. 9 and Note 12) to see for yourself.

The Mellor quotation (sect. 2.1, para. 11) raises a more complex question. In addition to more ordinary phenomenal experiences, we sometimes have phenomenal experience of *relations* among some of our phenomenal experiences, relations that involve "co-consciousness" of the experiences. Choosing a nontemporal example, we may experience one pain as more intense than another pain in part by experiencing both pains. The more complex question is this: Can we be conscious of the intensity relation between two pains (including co-consciousness of the two pains) without the two pains themselves having the intensity relation? Of course, I can judge that the headache was more intense than the backache without ever being co-conscious at all of the two pains. But the issue here is not one of judgments, but of experience of relations among experiences. Mellor takes a stand on this issue for the special case of temporal relations. D & K treat this view with scorn, even linking it to "the invited conclusion that all perceptions of temporal order must be accomplished in a single place" (sect. 2.1, para. 15). Their tacit rejection of phenomenal consciousness influences their interpretation of those who accept it, and so they see Cartesian materialism lurking under every bush.

Cartesian modularism. In concentrating on Cartesian materialism, a view that D & K concede probably no one holds explicitly, they ignore a far more interesting view that is a genuine object of contention. The more interesting view is Cartesian modularism, the claim that there is a *system* for phenomenal consciousness. In talking of Cartesian materialism, D & K usually interpret "place" as *physical place*, but occasionally they talk as if they want "place" to cover *functional* place as well (sect. 1.1, para. 8). They don't elaborate, but on the functional understanding of Cartesian materialism it would be Cartesian modularism, the view that all conscious events occur in a single system. Then, however, the neurophysiological evidence D & K give against there being an actual place in the brain where all phenomenally conscious events occur would have no relevance. What does spatio-temporal spread have to do with Cartesian modularism?

I explore Cartesian modularism elsewhere (Block, forthcoming), but briefly, it is an interesting fact that the following co-occur:

1. Phenomenally conscious events.
2. Events that have access to inferentially promiscuous reasoning processes.
3. Events that guide action and speech.

This co-occurrence, along with many other considerations, suggests a model in which there is a single system or group of closely connected systems that subserve these functions. Schacter (1989) puts forward such a model; critics of phenomenal consciousness should take it seriously.[2]

NOTES

1. See Shoemaker (1981) on the inverted spectrum; Block (1978) on absent qualia; Nagel (1974) and Levine (1983) on the explanatory gap. See van Gulick (forthcoming) and Flanagan (1991) for general treatments. Note that I do not share McGinn's (1991) skepticism about the scientific investigation of phenomenal consciousness (though in other respects our positions are similar). Note also that although the concept of consciousness is not identical to any functional or information-processing concept, that does not preclude an identification of the property of consciousness (as opposed to the concept) from being identified with a functional or information-processing property. The property/concept distinction cannot be clarified here. See Loar (1990).

2. I am grateful to Michael Antony, Paul Boghossian, Stevan Harnad, and Stephen White for comments on an earlier draft.

What is consciousness for, anyway?

Bruce Bridgeman

Program in Experimental Psychology, University of California at Santa Cruz, Santa Cruz, CA 95064

Electronic mail: *psy160@ucscc.ucsc.edu*

Dennett & Kinsbourne (D & K) take consciousness for granted and go on to describe some of its characteristics. In a passage reminiscent of the Bible's "In the beginning, when God created . . . ," they start section 1.1 with "Where there is a conscious mind. . . ." Although their analysis is certainly productive, it should be grounded in a functional perspective of the conscious faculty as a neurological system like any other, with a job to do and a way of doing it. This commentary will attempt to provide that grounding. The result will come out somewhat backwards, with the conscious mind emerging as a result of the operation of other systems.

The starting point is evolutionary theory. We know that because of their complexity the neurological systems that support consciousness must have evolved over a very long time and were retained because they had a function. The basic function (elaborated from Norman & Shallice 1980 and Bridgeman 1988) originates from the need of more advanced animals to separate behavior from the immediate exigencies of the environment.

An organism that merely reacts to sensory information has no need for consciousness – it simply does what the environment demands and its psychology is one giant transfer function. As soon as more than one plan drives behavior, however, there must be an internal rather than an external trigger for action. Along with this must come a planning mechanism that makes

plans, stores them in memory, and keeps track of where the execution of each one stands.

I suggest that the escape from stimulus-response behavior makes consciousness necessary, a keeping track in memory of internal rather than external controls on behavior. It is this continuous plan-monitoring function, and nothing more or less, that we define as consciousness. Thus consciousness is emergent from the process of driving behavior from internally held plans. Consciousness is not an object in itself, but a side effect of other neurological operations. In defining consciousness in this way we also redefine psychology, from a study of stimulus-response contingencies to a study of the plans that drive behavior.

There is no room for a "Cartesian Theater" in this conception, but the function of consciousness clearly fits with a parallel "multiple-drafts" notion, for many plans exist simultaneously (everything from short-term plans such as cooking dinner to long-term ones such as earning a Ph.D.). The control of behavior is a giant juggling act. Existing plans are evaluated along with incoming sensory information and the execution of one plan wins out. At the same time, new plans and new subparts of existing plans are created.

This interpretation of the function of a conscious mind has consequences in several illusions where the differences between consciousness and reality are particularly evident. One such consequence is the illusion of the knife-edge of time discussed by D & K. There is no need to micromanage temporal relationships in a range where time distortions in the incoming sensory channels and the motor apparatus begin to become significant. But the concept of the knife-edge precision of definition is necessary to assign a temporal order, however arbitrary, to events and actions. As D & K point out, the temporal tag is not itself a temporal event; it is useful in the planning and plan-executing processes, not in the real-time operation of the brain.

A second illusion, not discussed by D & K but illuminating in this context, is the feeling that the visual field presents a detailed and veridical representation of the surrounding world. Everyone shares an almost irresistible introspection that the visual world present in consciousness is, for example, in full color and sharp focus. Yet we know that reasonable color coding exists only in the central 30° or so of the retinal image, and that sharpest focus and high-acuity imaging occupy only a tiny region in the fovea. We see not the retinal image, but some idealized combination of sensory information, memory, and assumption combined so seamlessly that we are unaware that most of what we perceive isn't actually available in the retinal signal. It is just this composite that is useful in making decisions based on visual information – the immediately present visual image is just a processing stage, a small part of the available information.

A third illusion is the distal reference that characterizes both sensory and motor operations. We perceive objects in the world, not in the eye or ear, and we feel objects, not deformations on the skin. Awareness arises only where it is functionally advantageous, at a level of coding where sensory and motor processes are coded as common, distally oriented events (Prinz 1991). A wealth of empirical work (summarized by Bridgeman 1990) is now available that distinguishes what processing is available in consciousness and what is not.

It may seem ironic that the processes of creating plans, accessing memory, and keeping track of everything are themselves unconscious, that we are unaware in consciousness of the functions that support consciousness. But if we think of consciousness as a result of planning capabilities, not as a system in itself that must be modeled, there is no reason why consciousness should appear at a mechanism level and there is no mechanism to make that possible. Consciousness is not a monitor of mental life but a result of mental operations separated from the immediate sensorimotor world.

Experiential facts?

Andy Clark

School of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton BN1 9QH, England

Electronic mail: *andycl@cogs.sussex.ac.uk*

In this timely piece, Dennett & Kinsbourne (D & K) attack what they term the image of the Cartesian Theater. The key move in their argument is to insist that just as judgments about, for example, redness do not require corresponding brain events that are actually red, so, too, judgments about temporal sequence need not involve the construction, somewhere in the brain, of an actual sequence of brain events that are temporally related in just the same way as the events which figure in the judgment (see e.g., sect. 1.1, para. 6; sect. 2.1, para. 6). To suppose that judgments of sequence must depend on the absolute sequence of brain events at some (functional) point where "it all comes together" is a mistake that, it is claimed, can make a variety of phenomena involving subjective judgments of temporality seem needlessly anomalous. I believe D & K are right to reject this strong image of the Cartesian Theater, but their argument goes further, for they then go on to deny the distinction between:

a. cases where an agent actually has a subjective experience of some event (like seeing a woman without glasses) but later comes to judge that the woman wore glasses all along (the "Orwellian" story – sect. 2.2., para. 7), and

b. cases where the agent's original experience was of seeing a woman with glasses (even though the woman in question was not wearing any), and this experience is accurately recalled in the later judgment that she had glasses on (the "Stalinesque" story, sect. 2.2, para. 8).

It is not obvious (to me at least) why this distinction needs to be denied. For such a denial is not *forced* upon us when we give up the image of the Cartesian Theater. We may agree that later judgments of the temporal order of events need not be grounded in a kind of action replay in which we construct brain events of matching temporality. But we may *still* believe that there were facts about the immediate contents of conscious experience so that it can (for example) simply be true (or false) that at time *t*, you had the experience of seeing a woman without glasses. Such conscious states may surely form an absolute temporal sequence independent of the agent's propensity to later judge that given conscious states did or did not occur. And this is, on the face of it, all that is needed to justify the Orwellian/Stalinesque distinction. In short, I don't see why recognising the errors D & K point out undermines the idea of an absolute timing of conscious contents or of absolute facts about conscious contents. Such facts could be quite independent of our later judgments, and be facts nonetheless.

D & K's position is somewhat clarified by the example of metacontrast. Here, faced with the question "Did conscious perception of a disc occur?" they decline to answer, saying that "information about the disc was briefly in a functional position to contribute to a later report, but this state lapsed" (sect. 2.2, para. 29). The question of whether the disc perception was ever conscious is one that D & K claim is opaque to both the agent and to any outside observers. Probe the agent at different times and you will get different answers. Both Stalinesque and Orwellian stories are, it seems, "consistent with whatever the subject says, or thinks, or remembers" (sect. 2.2, para. 36). But this is surely only true if "thinks" here means "later judges to be the case." If we believe that there are facts about conscious contents and that such facts are in principle independent of later verbal reports, the distinction can be maintained. Perhaps it is the idea of facts about conscious mental experiences independent of facts about later verbal judgments that D & K really seek to displace?

I suspect that this is indeed the case and that D & K really

want to cast doubt on the very idea of an experiential fact – a fact about the content of conscious experience at a given moment. The arguments concerning temporal and spatial smear (sect. 1.1) suggest that they wish to reject the very idea of a single conscious observer as a locus of experiential facts. But nothing in the explicit argument seems to justify this radical conclusion. We could grant that a variety of brain states (spatially distributed) could be implicated in the construction of immediate conscious contents and yet still discover that there is some functional property (e.g., of synchrony of neural activity in certain regions), which is both necessary for a content to become consciously known and yields an absolute temporal order of experiences (with specific contents) – an order that need not, however, be preserved in later judgments about the order. I cannot see that this possibility is ruled out by anything that Dennett and Kinsbourne tell us.

To sum up, the move from the (proper and important) rejection of the idea that judgments of temporality require a matching temporal sequence of brain events to the denial of the Orwellian/Stalinesque distinction looks problematic. The transition could be oiled by some radical views about the nonexistence of experiential facts or the relation between conscious content and verbal report. If there is indeed such a hidden agenda, it should come on stage for the curtain calls.

The selfless consciousness

Antonio R. Damasio

Department of Neurology, University of Iowa College of Medicine, Iowa City, IA 52240

Electronic mail: *cmdardpg@uiamvs.bitnet*

I enjoyed reading Dennett & Kinsbourne's (D & K's) target article. It provides interesting ammunition against the intuition that consciousness depends on a single brain locus where multifarious information comes together, in spatial and temporal terms. This notion, which D & K modestly refer to as a "prevailing view," is far worse than that: It informs virtually all research on mind and brain, explicitly or implicitly, and is certainly the common sense concept of the nonscientist and nonphilosopher in the street. My other comments are as follows:

(1) The evidence presented by D & K draws on cognitive science and is damaging enough to the Cartesian Theater model. I think, however, that D & K could have made an even stronger case by using evidence from experimental neuroanatomy, neurophysiology, neuropsychology, and computational neuroscience. Elsewhere I have analyzed part of that evidence to construct an argument against one or even a few "integrative" brain sites (Damasio 1989a; 1989b; 1990). For instance, there is no neuroanatomical structure in the cerebral cortex to which signals from all the sensory modalities that may be represented in our experience can converge, spatially and temporally. The entorhinal cortex and the hippocampus might be candidates for that sort of "integrative" role but they do not pass the necessary anatomical tests. Also, we know for certain that they cannot do the job because patients in whom such structures are destroyed bilaterally (e.g., patients Boswell [see Damasio et al. 1989] and H. M. [see Corkin 1984]) do not have a disturbance of consciousness in the sense discussed by D & K. (It can be argued that Boswell's highest level of self-consciousness is not intact since he cannot access a large body of unique memories from his past, but it is clear that he deals quite self-consciously and appropriately with the universe, at categorical level). The prefrontal cortex, another region associated with consciousness in the minds of most people that have ever thought about the brain, is an even less adequate candidate than the entorhinal cortex for the "integrative" locus underlying a Cartesian Theater. It provides many anchor points for signals hailing from various sensory streams and from the motor system, but there is no single site to which "representations" can cohere spatially and temporally. Extensive bilateral ablation of prefrontal cortices in humans does not preclude basic consciousness, although, again, we have argued that the highest levels of self-consciousness are not possible without these structures.

(2) I had two problems with D & K's proposal. The first and most important is that the rejection of *one* biologically impossible Cartesian Theater does not amount to rejecting the sense of *one* self doing the experiencing. There are, without a doubt, neural systems whose operation generates the sense of self, and on the basis of which we construct the false intuition that *there is one brain site where experience happens.* A satisfactory model of consciousness should indicate how the dis-integrated fragments operate to produce the integrated self. My impression is that the Multiple Drafts model is part of an alternative to the Cartesian Theater model but not a complete one. I would suggest that there are two necessary functions missing from the Multiple Drafts model without which I cannot fathom how consciousness, illusory phenomenon that it may be, will emerge. The first function is the sustained updating of critical sets of knowledge of the individual doing the experiencing. The sets encompass both taxonomically categorical levels of knowledge ("supraordinate" and "basic object"), as well as unique level ("subordinate" and autobiographical). The updated knowledge refers not only to the past but also to the future, that is, to memories of intended actions and plans. The second function is the sustained monitoring of somatic states of the experiencer, to include both visceral and musculoskeletal sectors of the organism. I suspect that the updating of previously acquired knowledge is implied in the Multiple Drafts hypothesis, but I saw no reference to the possible role of somatic states. I do not believe consciousness is possible without having something like the multiple drafts of D & K *referred to the somatic base of the experiencer.* Only awareness, in the sense used by Crick and Koch (1990), might be possible without a somatic reference.

I am persuaded that the multiple drafts mechanism alone will produce a selfless, disembodied consciousness. Incidentally, selfless consciousness can *almost* happen in some circumstances. An example is anosognosia, a neurological condition caused by extensive parietal and frontal damage in the human right hemisphere. The patients are unable to monitor their somatic states comprehensively and become unconcerned with their medical problems and with their future implications. They can give evidence that some externally generated representations of their own body do not pertain to themselves. I usually teach about this condition by stating that the lesion has "chipped part of consciousness away," that many percepts and thoughts of these patients are no longer referred to their bodies (for a similar perspective on the neurobiological basis of the self, see Merleau Ponty or, more recently, Edelman, 1989). Deep level meditation is another circumstance in which consciousness loses itself, so to speak, and eventually dissolves (this can be achieved in certain forms of Buddhist meditation).

(3) My second problem has to do with the degree of dissolution of the Cartesian Theater in D & K's alternative. How dis-integrated, neurally speaking, need dis-integration be? I agree that there is no single Cartesian Theater, but I suspect that there may be *many* such theaters, or, to use my own metaphor, many stages on which relatively coherent drafts of ongoing neural activity play out, at slightly different times. The point here is that it is plausible that some components of our experience actually depend on a local integration of neural sets of activity. For example, under certain circumstances this might happen in primary sensory cortices as a result of synchronization generated by feedback.

My closing comment is about the connection between consciousness and the timing of neural events. This is an important issue and is finally receiving the attention it deserves. That time can provide the illusion of a single place has been proposed by

several investigators (Crick 1984; Damasio 1989a; 1989b; 1990; Edelman 1989; von der Malsburg 1987). There is some preliminary evidence that the trick may actually work (Gray et al. 1989).

The distributed pineal gland

Martha J. Farah

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213

Electronic mail: *farah@psy.cmu.edu*

Dennett & Kinsbourne's (D & K's) discussion of "the where and when of consciousness in the brain" is an example of the growing trend toward using neuroscientific data to address questions about consciousness. Such projects are often informative and worthwhile, but emphatically not for what they tell us about consciousness.

What do D & K establish in this target article? They show us that the representation of a given event or sequence of events may be distributed across brain regions and over time, and that the representation of temporal order need not be the temporal order of representations. Once we have accepted these very reasonable points, certain empirical findings that might have seemed paradoxical before seem unremarkable. This may be a useful contribution to the psychology of time perception, but it tells us nothing about consciousness. Their analysis would apply equally well to the nonconscious representations of an appropriately programmed PC. The constant reference to consciousness, rather than to time information or time representation, is gratuitous.

Similar complaints can be made about the recent work on consciousness in visual perception. Parietal-lobe-damaged patients with an attentional impairment known as "extinction" may fail to identify stimuli presented in the affected region of space, but they can nevertheless classify those stimuli as the same as, or different from, other stimuli (Volpe et al. 1979). This has been taken as a demonstration that the parietal lobe is needed for visual percepts to reach consciousness. However, it can also be explained without invoking consciousness, by the far less sensational hypothesis that the parietal lobe plays a role in visual information processing, and that a less precise visual representation is needed to decide whether a stimulus is the same as or different from another one, compared to identifying it uniquely (Farah et al. 1991). Prosopagnosic patients (unable to identify faces) seem to respond appropriately to previously familiar faces in certain indirect ways. For example, they can learn to associate faces with their correct names faster than with incorrect names, suggesting to some authors that the problem in such cases is not in face recognition per se, but in conveying the products of face recognition to conscious awareness (e.g., De Haan et al. 1987). Again, these data can be understood more simply and without any need for the concept of consciousness by the hypothesis that face recognition is impaired in these patients and that the tests of "unconscious" face recognition, such as savings in relearning, are simply more sensitive tests to detecting residual functioning of a damaged visual system (Wallace & Farah, in press).

The dissociations among different kinds of visual abilities after brain damage are interesting for what they tell us about graded information processing in the brain and about the ability of partial or low-quality information to support certain computations and not others. They have not, so far, shed any light on the mechanisms of consciousness. Similarly, D & K's collection of puzzling phenomena in time perception highlights the distributed nature of time representation and the importance of distinguishing the sequence of representations from the representation of the sequence. But, despite the subtitle of their target article and the numerous tantalizing references to consciousness in their text, D & K offer no insights at all about consciousness. Take the most eliminative and unsexy definition of the word "consciousness" you like, for example, control mechanisms, or accessibility to speech – D & K have still told us nothing about it. The most that can be said is that they clarify certain properties of a type of information processing (the information processing underlying time perception) that could be conscious or unconscious and that happens in at least some cases to be conscious.

The fact that D & K's ideas apply to time perception rather than vision or some other perceptual modality may reinforce the illusion of their relevance to consciousness. For reasons that cannot be discussed in the space of a brief commentary, time perception has long been associated with conscious awareness by psychologists (e.g., Ornstein 1977) and philosophers (e.g., Bergson 1910). If D & K had written a similar article accounting for anomalies in color perception rather than time perception, entitled "Color and the Observer," we would not be so tempted to look for insights into consciousness. [See Thompson et al.: "Ways of Coloring" *BBS* 15(1) 1992.]

In sum, as a claim about the computational and neural mechanisms of time perception, Dennett & Kinsbourne's idea about representations distributed in space and time seems fine. But as a claim about "the where and when of consciousness," it misses the mark just as widely as Descartes' theory of the pineal gland.

The Cartesian Theater stance

Bruce Glymour, Rick Grush, Valerie Gray Hardcastle, Brian Keeley, Joe Ramsey, Oron Shagrir, and Ellen Watson

Department of Philosophy, University of California, San Diego, La Jolla, CA 92093-0302

Electronic mail: *ga1043@sdcc6.ucsd.edu*

Even though we do not reject Dennett & Kinsbourne's (D & K's) conclusion, we have several difficulties with their arguments for the distinction between a Cartesian Theater model and a Multiple Drafts model. D & K argue that the Cartesian Theater model of Cartesian materialism fails because it requires that mental content cross a unique experiential finishing line in order to become conscious experience, that temporal content be carried only by temporal properties of representations, and that there be an empirically unavailable distinction between Stalinesque and Orwellian revision; they claim that the Multiple Drafts model succeeds on all three counts. However, Cartesian materialism need not fall prey to their criticisms; the Multiple Drafts model becomes either a version of a Cartesian Theater model or an eliminativist theory.

Cartesian materialism does presuppose a finishing line, because it assumes a single determinate stream of consciousness. Cartesian Theater models do require that there are facts about when a represented content becomes part of the stream. However, they do not require that the activity of one particular area in the brain be correlated with consciousness – there may be many such areas. Grounds for rejecting a unique neurobiological finishing line are therefore not sufficient for rejecting Cartesian Theater models. One should instead reject the idea of a single "film" whose observation is equated with consciousness.

Moreover, the Cartesian Theater model, just like the Multiple Drafts model, allows the representation of temporal content by nontemporal features of the representations. What the Cartesian Theater requires is only that whenever a representation does reach a neurobiological finishing line, the content of the representation be experienced at that time, regardless of how broad or "smeared" that finishing line may be. We suspect that D & K mean to deny that an experienced temporal sequence is identical to the sequence of the individual experiences that compose it, but to deny that would be a mistake. Suppose a

temporal sequence is composed of experiences E1, E2, E3, in that order. Either the content of the three individual experiences is experienced in that order, too, or that temporal sequence is not experienced at all. The issue might be that the sequence of experiences represents a temporal fact about three vents e1, e2, and e3. E1, E2, E3 could mean that the events represented occurred in the order e2, e1, e3. This is a perfectly possible state of affairs the Cartesian Theater model does not rule out. However, this is not an example of the order of experiences differing from the experienced order; to believe that would be to confuse the order of experiences with the order of events represented by the content and order of the experiences.

Finally, D & K present the Cartesian Theater model as landing us in the empirically unresolvable Stalinesque/ Orwellian pickle, while the Multiple Drafts model obviates the need for the distinction. They conclude that there is no fact about whether any revision is Stalinesque or Orwellian and so the Multiple Drafts model is preferable. But even if D & K's epistemological verificationism were acceptable, there is still a way to answer questions about types of revision (at least in some cases). D & K accept verbal reports as indicators of conscious experience; however, they do not argue that these reports are the only such indicators, nor that they are not correlated with neurobiological activity. At first blush, neurobiological evidence might well prove a robust indicator of consciousness. For example, if we restrict ourselves to measuring verbal reports and gross bodily movement, there is no observable difference – in either first or third person – between drugs that render subjects unconscious and oblivious to painful stimuli (a Stalinesque experience) and drugs that do not prevent pain from being experienced but that temporarily paralyze the subjects and keep the pain from being recalled (an Orwellian experience). But we do have access to more than mere verbal reports. For example, anesthesiologists know that alleged analgesics produce EEG waves similar to those found in unconscious sleeping subjects, while alleged paralytic amnesics produce EEG's similar to alert subjects. (For similar notions, see Kulli & Koch 1991). Once again, we are left without a good reason to prefer the Multiple Drafts model to the Cartesian Theater model.

Moreover, as long as the Multiple Drafts model includes the notion of consciousness as an observer and interpreter of some "draft," it seems reasonably close to a version of Cartesian materialism in which the "theater" shifts around in the brain, since it still requires a privileged and unique temporal "finishing line" for each content probed in the multiple drafts. If there were no such line and no definitive narrative of conscious experience, then it would not matter whether any contentful state were probed at all. D & K suppose contentful states can be causally efficacious without leaving any trace in consciousness, and their Multiple Drafts model would support the lack of functional significance between conscious and nonconscious states. But in this case, a Multiple Drafts model would not preserve the role for consciousness as an interpreter, which D & K explicitly include. And if they do not go on to draw this more radical conclusion, D & K have not really departed from the Cartesian Theater.

Nothing is instantaneous, even in sensation

Robert A. M. Gregson

Department of Psychology, Australian National University, Canberra, A.C.T. 2601, Australia

Electronic mail: *rag655@csc2.anu.edu.au*

Let us start with processing time; a stimulus input that may be virtually instantaneous to the external observer, like a Dirac delta function, is converted into one or more temporally extended representations in the neural mass of the brain. For sensations, at least two pathways appear to be implicated (Mishkin & Appenzeller 1987) and each can form closed loops. This initial activity can be short lived or turned into limit cycling, which at lower phyletic levels may then perseverate for hours, even days. The ostensive "finish line" is a consequence, then, of tapping the extended representation and converting the sample of such activity into phenomenology, which in turn becomes a verbal or motor response to the outside observer. Or perhaps the phenomenology is bypassed completely, if we create a demand situation that induces that bypassing.

The two externally observable events can be at almost exact points in time, but unless they were coincident, which they never are, any intervening events need not themselves be instantaneous. Theory that assumes they are temporally extended can be constructed and can fit some qualitative facts in psychophysics (Gregson 1988; 1991). The response, a sensation, can itself be quite long and apparently slowly evolving, as is the case in gustation. I see no reason to be depressed if my consciousness production mechanisms lags behind my other more efficient brain processes; quite the contrary, I could argue that any organisms built that way might have evolutionary advantages or that we have inherited something that was advantageous in our primitive ancestors and now is mostly innocuous.

From such a position it seems we could rule out the Cartesian Theatre and only consider the Multiple Drafts. The question that interests me is where in all this do we get our basis for reporting when events arise in our consciousness, and how do we estimate the passing of time between two or more successive events? The distinction between time represented and time of reporting seems to suggest that when we initiate a brain process, we also set running in parallel a clock which puts markers on the record. The coarser the clock's resolution in real time, the more fuzzy the representations of successive process stages. Now, there are three alternatives which dangle themselves before our metaphorical eyes:

i. each process initiates its own clock, autonomous from all other clocks and the last register on that specific clock is when the process is terminated, vacated, or overwritten, or

ii. there is a common master clock and each process picks up a starting label (the postmark) and runs at the same rate in real time thereafter, or

iii. there aren't any clocks but the processes have paths in real time that are characterized by intrinsic and inescapable periodicities.

Such quasiperiodicities are typical of nonlinear dynamics, which we know do occur in the brain. This third position is a strong version of the first, it says that all processes are autonomous and clocks are not needed because every process is a time series which is generated by an attractor's dynamics. The autonomy part of this is not strongly asserted; we can have dynamics that, as it were, leak into one another. This third position is the one to which I have subscribed (Gregson 1988; 1991) but it leaves us with the problem of how we can ever talk about the simultaneity of steps in two contemporaneous processes. If it is felt that we must have in the brain a third process, which acts as a monitor to the first two, then we are half way back to the Cartesian Theatre with one great monitor, like Bentham's panopticon. This is dubious, for even if there are no special continuously running processes always accessible to act as monitors, each coupling of two or more processes should still be able to function as a single dynamic with a single time marker, but with three coupled processes (or even three degrees of freedom) the process becomes *sometimes* chaotic (Nicolis 1986). Coupled dynamics typically exhibit a master to slave relationship, one inducing a forcing function onto the periodicity and hence the timing of the other (Haken 1985). It is possible to write dynamics that are only open to input at periodic points (Zak 1991) and would therefore only collect input from another process at those "windows," even if the processes are continuously coupled.

Suppose we have two processes, a and b, with values V_{ai}, V_{bi}, $i = 1,2, \ldots$, starting at times t_{a°, t_{b°, $t_{a^\circ} < t_{b^\circ}$. Let a have a quasiperiodicity θ_a and b have θ_b; both can interact only when their own dynamics admit input (in technical terms, when the Lipschitz condition is violated). Then a is first influenced by b only at the first synchrony of θ_a and θ_b, and when this occurs depends on all of $\{t_{b^\circ} - t_{a^\circ}, \theta_a, \theta_b\}$. If a tries to "deduce" from its first input, V_{b1} from b, whether $t_{b^\circ} - t_{a^\circ}$ is positive or negative, it has a most uncertain basis on which to do just that, even if both processes have, for a while, *monotonically increasing values* that are intermittently revealed through their windows.

I don't think we need all the delightful illustrations the authors have provided, for if the percepts are based on the consequences of nonlinear dynamics we should expect and even predict curiosities in any comparative responses involving two channels of input when those responses are constrained to happen quickly, and particularly when the internal representations of inputs rise and fall at different rates. The reported paradoxes will happen in sensation mechanisms, even if those mechanisms have error-free internal memories of the outputs of their lower level processes such as V_a, V_b.

From my perspective, neither the Stalinesque nor the Orwellian theories seem necessary, and certainly no projections backwards in time; we are not into parapsychology. Quite simply, coupled discontinuous nonlinear processes can create information that is misleading if there is an attempt to reconstruct, *at a fine-grained level*, the precise history of the past inputs, be it their past values or their past starting points.

I distrust examples about long-term memory errors in this context; for example, if I cannot recall (without access to my appointment book) whether I visited my bank manager or my dentist first on a Tuesday six weeks ago, this is not the same sort of failure as that occurring in phi-phenomena. There was no confusion on that Tuesday, even if both visits involved painful extraction, but each on that day had a clocktime label, and there was for a while recall of the journey from one office to another between the visits. Such vignettes are perhaps extended metaphors to illustrate Dennett & Kinsbourne's own theory, but they actually arise in such slow real time, with so much collateral input about clocktimes and sequence, that they are better eschewed.

Some mistakes about consciousness and their motivation

S. L. Hurley

University Lecturer and Tutorial Fellow in Philosophy, St. Edmund Hall, Oxford OX1 4AR, England

Consider the relationships between, and motivations for, the various mistakes Dennett & Kinsbourne (D & K) diagnose, which, as they rightly note, are strangely persistent. The first mistake, M1, is embodied in the metaphor of the Cartesian Theatre, and is labelled by D & K "Cartesian materialism," which they describe in terms of the idea of a centered locus of subjectivity in the brain where "it all comes together" into consciousness, and which, though material, takes over the role Descartes gave to the immaterial mind. The second mistake, M2, is the conflation of properties of representings, or vehicles of content, with properties represented, or of content itself. D & K are especially concerned with temporal properties, where the conflation is particularly tempting, but the conflation is mistaken quite generally.

These mistakes seem independent, in the (A) and (B) describe possible positions:

A. There might be a central processing unit in which the vehicles of all contents of consciousness were to be found, but which did not use properties of those vehicles to represent properties in content: did not use temporal sequence to represent temporal sequence, spatial properties to represent spatial properties, and so forth.

B. Conversely, there might be a noncentral, temporally and spatially distributed process that did use properties of vehicles to represent properties in content (however inefficient that might be). D & K in effect recognize that (B) is a possible position in section 3.2. I suggest that what is really at issue here is not the centrality in some privileged place of representation, but rather the tendency to conflate properties of vehicles of content with properties of their content, and that these issues are indeed in many respects independent (but see remarks on the unity of consciousness below). Sometimes in D & K's presentation their use of the theatre metaphor and talk of a central privileged place obscures the fact that the work of their argument is really being done by the exposure of M2 (as is the case with respect to M3 – see below). It is not clear why they regard M1 and M2 as necessarily associated or how the central place idea might explain the hold M2 has on us.

Why *do* we find it so hard to give the M2 conflation up? I don't sympathize with M2, but I want to speculate further about the source of its hold on us and to suggest that it has to do with both the unity and reflexivity of consciousness. We seem to want the relationship between vehicle and content to be an intrinsic one: This vehicle must, by its very intrinsic nature, bear this content. Moreover, it must do so transparently, so that the vehicle displays the content it bears to anyone who cares to look. If vehicles give themselves content in virtue of the copying of properties of vehicles into their content, then these desires seem to be met: Someone, some homunculus, can simply look at their intrinsic properties to see what their content is. And we're off on the regress. We get started on it because we cannot accept that properties of vehicles may carry their content, as it were, blindly, in virtue of some mere mapping or contextual relationship: Content then seems somehow too arbitrary, or too extrinsic and abstract to reflect the unity of self-conscious experience. How could various abstract functional properties of mere events amount to *this* specific unified field of experience, as I reflect on my experience when I look out the window? Perhaps M1, the central place idea, is connected with M2, the conflation, by reflection on the unity of consciousness: Perhaps out of a misguided urge to understand the unity of the contents of consciousness, we attribute features of unity to some central processing place where all the vehicles get together – the Cartesian Theatre. If we insist with D & K that relations of vehicles cannot be just bodily hauled across the vehicle/content distinction into content, and that properties of content cannot be projected back onto their vehicles, the unity of consciousness can be left looking very mysterious. Well, it *is* very mysterious; but the point to keep hold of here is that the unity of the collection of vehicles in the theatre provides only an illusion of understanding the unity of consciousness. Part of the hold M2 has on us may also stem from a certain view of self-consciousness: namely, that consciousness is a fully reflexive relation; whatever it is that is conscious must be fully self-conscious, conscious that it is conscious, so that the content of consciousness must reflect that which is conscious. Hence the tendency to project properties of "bearers" of contents of consciousness into the contents of consciousness: to assume, in the style of some of those D & K criticize, that the time of a representing must be represented in the content of the representing, and hence that the representing of times licenses inferences to the times of representing.

D & K also identify the mistake, M3, of thinking that there must always be a fact of the matter as to whether experience is revised by memory or tinkered with ab initio, whether the Orwellian or Stalinesque ploys are operative. They illustrate persuasively the way in which M3 depends in various cases on M2. Just because certain temporal distinctions must apply to vehicular processes, it doesn't follow that isomorphic distinc-

tions must apply within content, so that there must be a fact of the matter: was experience like *this*, then immediately altered with all traces of revisionary tinkering removed, or like *that* all along? In certain cases, D & K claim, there may be no difference from any possible point of view, including the first person, between these two possibilities, hence no fact of the matter.

Might M3 be motivated by anything other than M2? It might seem to be independently motivated by antiverificationism or by a kind of realism about experience. However, it's not at all clear that the fact-of-the-matter view that M3 identifies as mistaken can be defended by a sound antiverificationism or by realism about experience. To claim that a difference in the content of conscious experience must be a consciously experienceable difference is not to subscribe to verificationism, of which some may be tempted to accuse D & K. Nor is there an issue here about possible inability to manifest nevertheless real differences in experience: D & K's argument does not depend on requiring manifestations to observers of differences of experience, but will run entirely from the unexpressed, introspective first person viewpoint of content of experience. "Do the experiments on yourself and see," they can say. Since D & K's position on M3 does not need antirealism about experience, their view cannot be resisted by means of realism about experience. To object that the "observer" to whom manifestation is being required is oneself at a slightly later time is to assume the temporal atomism about consciousness that is D & K's target. But one can reject such atomism without thereby adopting antirealism about experience. If there are arguments independent of antirealism about experience for rejecting such atomism, such as arguments from what D & K call "temporal smear," then realists as much as antirealists must adjust their conception of the nature of consciousness – not merely their view of the possibility of manifesting it.

ACKNOWLEDGMENTS

I am grateful to the British Academy for supporting this work in the form of a Senior Research Readership. I am also grateful to Derek Parfit for discussion of these ideas.

The where in the brain determines the when in the mind

M. Jeannerod

Vision et Motricité, INSERM U94, 69500 Bron, France
Electronic mail: *Inpeu94@frsun12.bitnet*

There are situations in everyday life where actions in response to visual events are clearly dissociated from conscious experience of the same events. We respond first and become aware later. One example is when in driving a car we have to make a change in trajectory because of a sudden obstacle: We consciously see the obstacle *after* we have avoided it (and only then do we experience retrospective fear!). Castiello et al. (1991) designed a series of experiments in which they measured this temporal dissociation. Suppose you are a subject in this experiment: You are instructed to reach with a hand for an object placed in front of you, as soon as it becomes illuminated. It will take you approximately 330 ms on average to start moving. Now suppose that, on a different set of trials, the experimenter's instructions are simply that you signal (with a vocal utterance: Tah!) at what instant you became aware of the illumination of the object. The vocal response will take 380 ms to appear. Finally, on still another set of trials, the instructions are to perform the two tasks at the same time. Not unexpectedly, the values for both the motor and the vocal reaction times will be found to be very close to those measured in the previous sets of trials, that is, the onset of hand movement aimed at the object will precede the vocal response signalling your awareness of its change in visual appearance by about 50 ms. This difference will not be noticeable to you; you will feel that your hand movement coincides with your perception of the illumination of the object.

Does this result mean that it takes approximately the same amount of time for the brain to generate a motor response to a visual event as it takes it to generate a subjective experience of that same event? What would happen if the motor reaction time were reduced? Would the time to awareness also become shorter, or would it remain the same? The experiment for answering these questions becomes a little more complicated: You are now facing three identical objects, each separated by about 10 cm. The instructions are as in the earlier version of the experiment: reach for the object that is illuminated and signal vocally the time at which you become aware of it. On most trials, the central object alone is illuminated. On some occasions, however, the light that illuminates the central object is suddenly shifted to one of the other two. This shift is triggered exactly at the time of onset of the hand movement. In this event, according to the instructions you received, you will have to correct the direction of your hand movement in order to track the second illuminated object, and to give a second vocal signal to indicate the time when you became aware of the shift in illumination. The first sign of correction of the hand trajectory will appear early (about 100 ms) following the shift in illumination. By contrast, the vocal utterance corresponding to this same event will come much later, on the order of 300 ms after the beginning of the change in movement trajectory. Your subjective report will be in accordance with this temporal dissociation between the two responses: You will report that you saw the light jumping from the first to the second object near the end of your movement, just at the time you were about to take the object (sometimes even *after* you took it!).

The clearest effect observed in this series of experiments is that the time to awareness of a visual event, as inferred from the vocal response, retains a relatively constant value across different conditions (for details and control experiments, see Castiello et al. 1991). Under "normal" circumstances (i.e., where no time pressure is imposed on performing the task) this value is roughly compatible with the duration of motor reaction times. The consequence of this compatibility is that when you make a movement toward an object, you become aware of this object near the time when the movement starts, or shortly after it has started. Hence the apparent consistency between our actions and the flow of our subjective experience. This consistency does not seem to be affected by small differences in timing (up to 50 ms), but it breaks down when the difference increases, for instance, when the motor reaction time becomes shorter under conditions of time pressure, such as avoiding sudden obstacles or tracking unexpected object displacements.

It would be tempting to speculate about the long duration of our motor reaction times in normal, unperturbed, conditions. Long reaction times might have the function of keeping our subjective experience in register with our actions. Imagine what our life would be if the above temporal dissociation were the usual case, and if our awareness of the external events were systematically delayed relative to our actions in response to these events! More seriously, this dissociation between motor responses and subjective experience, when it happens, as well as the more usual synchrony between the two, reflect the constraints imposed by brain circuitry during the processing of neural information. Different aspects of the same event are processed in different pathways. Pathological cases give many examples of such dissociations created by lesions affecting one of those pathways while sparing the others.

I do not fully agree with Dennett & Kinsbourne when they conclude that "the representation of sequence in the stream of consciousness is a product of the brain's interpretive processes, not a direct reflection of the sequence of events making up those

processes." I can certainly accept that the experienced sequence of events does not reflect a sequence of processes, but this is only because the neural processing (of motor responses, of object recognition, etc.) is not sequential and because it occurs in parallel pathways (this is indeed what the Multiple Drafts model is about). It happens that these pathways are tuned to processing information at about the same rate, or, more correctly, that the rate of processing is globally constrained by the slowness of the representational pathway. But the impression of simultaneity that one gets is not an illusion or a post-hoc reconstruction. The apparent synchrony (or the apparent dissynchrony) of awareness with other events (or the fact that the awareness of a given event does or does not occur at all, are directly related to the amount and the duration of neural processing needed to achieve conscious experience. As shown by the above experiments, consciousness is not immediate, it takes time to appear (see Jeannerod, 1990, for review). Adequate timing of neuronal activity in different brain areas is a critical condition for achieving subjective temporal consistency between external events.

Models of conscious timing and the experimental evidence

Benjamin Libet

Department of Physiology, University of California, San Francisco, CA 94143-0444

1. Cartesian Theater versus Multiple Drafts models. The problem of how the brain "binds" various separately discriminated contents into a unified subjective experience is an old one (e.g., Sherrington 1940). D & K argue that unifying processes are themselves distributed, and as a result, the "point of view of the observer is spatially smeared" (sect. 1.1). Such a spatially (and also, they claim, temporally) smeared point of view will hardly solve the problem of a unified conscious experience. If the "Multiple Drafts model avoids the tempting mistake of supposing that there must be a single narrative . . . that represents the *actual* stream of consciousness of the subject" (sect. 1.1, para. 9), how does the Multiple Drafts model explain the actual experience of a single narrative we all have?

D & K's reason for believing that "slow transmission and computation speeds of neurons" would create a temporal smear of up to several hundred milliseconds is not a valid representation of the neurophysiological evidence. Transmission is fast enough to get messages around in milliseconds, or tens of msec, even when synaptic stations are involved. That there is nevertheless a neural delay of up to 500 msec for the development of a conscious experience is based on a requirement for suitable repetition of neuronal patterns, as experimentally demonstrated by Libet et al. (see Libet 1973; 1982; 1989b; Libet et al. 1991). D & K go on to propose, mistakenly, that "we need other principles to explain the ways in which *subjective temporal order* is composed, especially in cases in which the brain must cope with rapid sequences occurring at the limits of its powers of temporal resolution" (sect. 1.1, para. 6). Rapid cognitive and conative responses can be mediated by shorter-lasting activity patterns and can be made unconsciously, without or before the appearance of the *subjective* experience of an event (Libet 1965; Libet et al. 1991). D & K confuse the ability to make conscious discriminations of temporal intervals of, say, 50 msec with the question of *when* one becomes *aware* of the discrimination; that is, one may be able to discriminate a 50 msec interval and even act on that (in a reaction time test) but not be aware of the discrimination until some hundreds of msec later.

It seems clear to me that the various "temporal anomalies" of consciousness (sect. 1.2), which D & K claim are best explained by their Multiple Drafts model without the alleged erroneous and confusing implications of the Cartesian Theater, can be encompassed by either model. If D & K are proposing a serious scientific theory it is incumbent upon them to indicate experimental designs that could test and potentially falsify their proposal. Since no such tests are offered by them, the Multiple Drafts model is simply one special philosophical construction that competes with the others.

Incidentally, D & K erroneously state that Libet (1982; 1985a) "argued that these temporal anomalies are *proof* of the existence of an immaterial mind that interacts with the brain in a physically inexplicable fashion" (sect. 1.2, para. 1; emphasis mine). Libet has in fact said that the evidence and phenomena are compatible with any theory of the mind-brain relation (e.g., Libet 1985a, pp. 536, 563, 564; 1981, p. 196). My statement (Libet 1981, p. 196) that "the temporal discrepancy (between neural and subjective timing) creates relative difficulties for identity theory, but that these are not insurmountable" does not match D & K's ascription.

1.1. The representation of temporal properties versus the temporal properties of representations. Libet et al. (1979; Libet 1982; 1989b) had already not only made this distinction explicitly but they also obtained experimental evidence that directly demonstrated it (in the subjective appearance of sensory experience before the time of the adequate neural representation for the experience). The timing *content* of the sensory experience does not, however, appear in a "temporally sloppy way, anywhere in the brain" (sect. 2.1), as D & K would have us believe in their model. The experimental finding was that subjective timing of a sensory event is rather rigidly correlated with, though not necessarily identical to, that of the early primary cortical response to the signal, even though this response does not itself elicit the experience (Libet et al. 1979). In fact, in the absence of that early neural reference signal the subjective timing is not antedated and the sensory signal is subjectively experienced as actually having a substantial delay related to the delayed neural representation! Such a timing signal or "postmark" to instruct the content of a subjective experience was thus experimentally established by us long before Glynn (1990) suggested it was theoretically possible (see Libet, 1991a, reply to Glynn).

1.2. "Orwellian" and "Stalinesque" revisions (sect. 2.2). The so-called Orwellian postexperiential revision of memory or the Stalinesque preexperiential introduction of unconscious processes are both logically possible alternative explanations of some observed responses by a subject. They have in principle both been suggested earlier as possibilities (e.g., Libet et al. 1979; Libet 1978; 1981; 1985a), without applying the pejorative names of Stalin and Orwell to them.

Let us examine how D & K apply these arguments to the phenomenon of *metacontrast*, in which a second stimulus can inhibit or mask the reportable conscious experience of a preceding, first stimulus. They add that "both the Orwellian and the Stalinesque versions of the Cartesian Theater model can deftly account for *all* the data – not just the data we already have, but the data we can imagine getting in the future" (sect. 2.2, para. 19). Well, in our study of retroactive effects of a delayed cortical stimulus on the reportable awareness of a preceding skin stimulus, we found not only retroactive inhibition or masking but also retroactive *enhancement* of the conscious sensation of the initial skin stimulus (Libet 1978). Now, one might argue that our retroactive masking could be due to obliteration of memory for the skin sensation by the delayed cortical stimulus (although there is no evidence for such a process when one uses a highly localized cortical stimulus, and not a massive electroconvulsive shock to the whole brain). But in retroactive enhancement there is no memory loss at all; this finding therefore clearly supports the view that the content of the experience can be modulated before the experience appears. Of course, D & K may argue that the retroactive enhancement could be due to an "intensified"

memory rather than to a pre-experiential modification of content; but "intensification" of memory would be an ad hoc construction without any evidence for such a process.

In our more recent study, we have shown that a brief train of stimulus pulses (say 0.25 sec duration) in somatosensory thalamus can elicit correct detection of the signal by subjects even when they report feeling nothing and just guessing, whereas a longer stimulus train (about 0.5 sec or more) converts the responses *from* correct-with-no awareness *to* correct-with-awareness (Libet et al. 1991). Memory is not an issue in any of these trials; whether stimulus trains were short or long, subjects made their forced choice detection selections and their reports of awareness some seconds after stimulus delivery. Clearly, subjects had a memory of the brief stimulus, which they later correctly identified, but no reportable awareness. How would D & K's models deal with this unconscious detection of brief stimuli? Is there still no distinction between unconscious and conscious time factors?

2. *"The Libet controversies re-examined"* (sect. 3)

2.1. Technical criticisms. D & K takes it upon themselves to pronounce that Libet's experiments have "serious technical flaws" (sect. 3.1), that the experiments only "allegedly" show backwards referral, that "technical criticisms of his experiments and their interpretation raise doubts about the existence of the phenomena he claims to have discovered" (sect. 1.2, para. 2), etc. One would have expected thoughtful scholars to be more cautious about such pronouncements. D & K's referenced authority for all this is P. S. Churchland (1981a), who is a philosopher with minor experience in experimental neuroscience. Churchland's criticisms were replied to in detail by Libet (1981), for readers who wish to re-examine that controversy. There were some additional arguments about interpretation of some of the data by some of the commentators on my *BBS* target article on voluntary action to which I also replied (Libet 1985a). It should be noted that both of our experimental papers on the temporal anomalies had to pass rigorous peer review by the leading journal of neurology *Brain* (Libet et al. 1979; 1983a), whose editorial staff included distinguished experimental neuroscientists. These experimental findings have also received acceptance in international and national forums of neuroscience where they have been frequently presented since 1964. Interested readers should, in any case, check our original papers to satisfy themselves about the technical questions raised. When I challenged Dennett in a recent symposium to name specific difficulties with our findings, he raised that of regarding the subject's reported "clock time" as a valid indicator of the actual time of the subjects' conscious intention. The difficulty could arise in view of our own evidence for a delay of up to 0.5 sec before there is neuronal adequacy for the appearance of the visual experience of the clock-time. This difficulty had already been raised in commentaries by Latto (1985) and Rollman (1985) and replied to by Libet (1985a), to which readers are referred.

D & K specifically state as technical indictments that Libet's experiments "involved very few subjects, were inadequately controlled, and have not been replicated" (sect. 3.1, para. 1). In fact *hundreds* of subjects have been studied by us over the years. As indicated in our paper on subjective referral (Libet et al. 1979), results in a large number of subjects contributed to the development of the argument, but only in a relatively small number of patients could we run sufficiently elaborate and complete studies amenable to rigorous statistical analysis. However, this was sufficient to make the findings not only qualitatively but statistically convincing. Since when is the precise number of subjects necessarily a controlling factor in obtaining significant results? It is of course the amount and kind of data on each individual that is crucial. Clearly, the editors of *Brain*, with a professional background in dealing with research on human patients, regarded the data as sufficient for establishing our conclusions. They also found no difficulties with the controls, as detailed in the paper, something that scientific reviewers pay special attention to. But readers may check these points for themselves, in our original papers.

The additional charge, that our experiments "have not been replicated," displays a lack of understanding about such research as well as of awareness of some actual replications. To carry out the studies with intracranial electrodes in fully *awake*, cooperative human subjects requires the availability of special patients as well as an interested neurosurgeon willing to make the proper effort needed for a basic research study that has no direct clinical implications at all; it also requires a neuroscientist (like myself) with the dogged determination and patience to pursue such a difficult study and with the ability to ask soluble questions amenable to workable experimental designs. It is therefore frustrating to be told that our basic experimental findings are questionable and flawed because others have not sweated it out through similar studies. Mind you, D & K do not cite any contradictory results. In fact, to the extent that others have attempted similar experiments, there have been some confirmations. The study on the timing of conscious intention in relation to onset of cerebral activity (readiness–potential) has been completely duplicated by Keller and Heckhausen (1990), although these authors proposed a different interpretation of our results (to which I have now submitted a reply, Libet 1991b). Further information of that study, which does not require intracranial electrodes, has been made by Howard Shevrin and a co-worker (unpublished). The inability of a single stimulus pulse to somatosensory cortex or to sensory (ventrobasal) thalamus to elicit any conscious sensory experience in a human subject (Libet et al. 1964; 1967) has been confirmed separately by the neurosurgeons Ronald Tasker and Robert G. Grossman (personal communications) and by Vahe Amassian et al. (who stimulated sensory cortex with an extracranial electrode); the more elaborate studies of full neural delay and subjective antedating have yet to be attempted by others.

2.2. Neural delay for sensory experience and subjective antedating. D & K mistakenly believe that our estimate of up to 500 msec, for achieving "neuronal adequacy" to elicit a conscious sensory experience, "is determined by seeing how late, following initial stimulation (to the skin) a direct cortical stimulation can interfere with the consciousness subsequently reported" (sect. 3.1, para. 5). The "up to 500 msec" value was in fact based upon (a) the stimulus duration required in the cerebral sensory pathway from skin to somatosensory cortex; (b) the requirement of late event-related-potentials in the cortical response to a skin stimulus; (c) the interval for retroactive effects of a delayed initial stimulus (the one mentioned by D & K); and (d) the discontinuous jump in reaction times, from 200–300 msec to 600–700 msec, when subjects are asked to slow their reaction time deliberately by only 100 msec or so (e.g., Libet 1981).

Later in section 3.1, D & K ask how Libet can deal with the possibility that there could have been a replacement of the veridical memory of an early conscious experience of the skin stimulus with a false memory, namely, that the skin sensation never occurred. (This alternative would be opposed to our proposal that our retroactive masking is evidence for interference with cortical processes that are developing the sensory experience during the hundreds of msec after the skin pulse). I have already replied to this possible alternative by noting that we observed retroactive *enhancement* as well as masking (Libet 1978; 1981); there is no loss of memory with enhancement. Of course, I cannot prevent D & K from inserting an Orwellian demon who will also change the memory of the nature of an experience so that the sensation is "falsely" remembered as being stronger than with no conditioning cortical stimulus; why this demon should revise memories in this specific way only when skin stimuli are followed by cortical stimuli is a question D & K will have to answer!

Since "falsifying the memory" of an experience is offered by D

& K as a general alternative explanation for instances of nonreportable awarenesses, how would they deal with the following experimental findings in our tests of the subjective referral hypothesis: Stimulation of medial lemniscus (LM) was set at an intensity that required a minimum 200 msec train of pulses to elicit any conscious sensation. If we reduced the train to 150 msec the subject reported feeling absolutely nothing. But if the 200 msec LM stimulus was paired with a single stimulus pulse to the skin, delivered at the *start* of the 200 msec LM stimulus, the subject reported that the LM-induced sensation started at the *same* time as the skin-induced sensation (Libet et al. 1979). In D & K's Orwellian model, the experience should have actually occurred at the beginning of the LM train (or at least at the same early time as that reported for the skin stimulus), rather than (as in our "Stalinesque" antedating model) at or after the end of the LM train and then subjectively referred back in time to the cortical response to the *first* of the LM stimulus pulses. If D & K's Orwellian model were a valid alternative, why does the subject report no sensation at all with a 150 msec train in LM? That is, if the experience actually appeared at the onset of the LM train, why should it "disappear" if the train lasts 150 msec but be reportable if the train is 200 msec? Our clear explanation for this paradox is that the primary evoked cortical response to even the first LM stimulus serves as a specific timing reference for subjective antedating of the experience, which cannot *actually* appear until the end of the 200 msec train. Also, how would Orwellian memory revision explain the *different* subjective timings for LM versus cortically induced sensations? Our postulate in fact also explains simply (a) why the sensation induced by a stimulus train to somatosensory cortex is *not* subjectively reported to coincide with a skin pulse, contrary to the case for LM, as the cortical stimulus does not generate the required primary evoked response; and (b) why a patient with unilateral destruction of the ascending specific pathway (that elicits the primary evoked response) subjectively times a stimulus to the affected hand as coming several hundred msec after one to the normal hand when both stimuli are simultaneous (Libet et al. 1979; see also Libet 1981)!

D & K also present (sect. 3.1) an extended criticism of our asking subjects to report their awarenesses unhurriedly, within a few seconds after each stimulus trial. They ask what is wrong with the vocal reaction time (R.T.) experiment of Churchland (1981a), who asked subjects to say "go" as soon as they were conscious of a skin stimulus. This issue was already replied to (Libet 1981), but this apparently needs repeating. The point is *not* that we can disprove Churchland's contention about what her subjects were reporting; it is rather that it is clearly *possible* for those subjects to yell "go" before they are actually aware of the stimulus and that, therefore, such responses cannot be taken to be valid primary indicators of the time of the experience. What do D & K (and Churchland) think subjects think they are doing in the usual R.T. experiment when asked to press a button (instead of yelling "go") as soon as they are conscious of the stimulus? Well, there is conclusive evidence that, regardless of what they think, subjects can press a button with exactly the same reaction time whether they are aware of the stimulus or not (Taylor & McCloskey 1990)!

Incidentally, the spurious comments related to "'absolute' time, as Libet would put it" (see end of sect. 3.1), are based on a misrepresentation of my position *against* being able to discern absolute subjective time (e.g., Libet, 1989a, in reply to a *BBS* Editorial Commentary).

2.3. "Libet's claims about the 'subjective delay' of consciousness of intention" (sect. 3.2). It's difficult to deal with D & K arguments on this (and other issues) as they are distributed among a variety of contexts in their target article. In an earlier section (sect. 1.2D) D & K argue that this temporal anomaly may appear to violate the principle that causes precede their effects. This would only be true if D & K want to insist that neural events that initiate a voluntary act must also immediately produce the conscious intention for that act. But the experimental evidence is precisely that the initiating neural events produce conscious intention, but only after a delay of about 350 msec; where is the violation of cause and effect if *conscious* intention is *not* in fact the *initiator* of a voluntary action?

D & K attribute to me a description of events leading to the appearance of conscious intention (sect. 3.2, para. 9) that seems patently irrelevant. Nowhere in our findings or interpretation is it necessary for the subject's task to include having to decide which of the "spot-representations marching past, having made their way slowly from the retina" should designate the moment of conscious intention. The visual signals from the "clock" move rapidly (not slowly) to their representations at the cortex in a matter of some tens of msec at most; but such intervals are not significant when we are finding a delay of 350 msec for reportable conscious intention to appear after onset of a cerebral volitionary process. Further, our finding is noncommittal about whether the representation of the conscious experience is one of "arrival at a point" or "a matter of representation exceeding some threshold of activation over the whole cortex or large parts thereof."

In Section 3.3, D & K state that "Libet's experiment created an artificial and difficult judgmental task that robbed the results of the hoped-for significance." The task was no more artificial than many other experimental designs and the subjects did not find it difficult after some practice. The Grey Walter experiment that D & K believe to be a superior one easily involved as artificial and difficult a task as ours. Grey Walter's use of the signal from motor cortex to activate the slide projector only showed that motor cortex becomes active before one's muscles move in a voluntary action; he did not determine quantitatively the subject's reportable time of conscious intention. Grey Walter's subjects evidently reported not being able to veto their own initiated button push. That is of course explainable by the observation that the "motor potential" at the motor cortex appears at about 50 msec before muscle activation (Deecke et al. 1976); this is too late for any conscious control to be able to intervene, as the final motor commands are already in process (McCloskey et al. 1983). In our study (Libet et al. 1983b) subjects were able to veto after the time of conscious intention (W) which appears 200 msec before muscle activation and well before the motor potential. D & K's attempt to explain the inability of Grey Walter's subject to veto, as a "natural interpretation for the brain to settle on," is a meaningless non-neural construction.

Toward an identity theory of consciousness

Dan Lloyd

Department of Philosophy, Trinity College, Hartford, CT 06106
Electronic mail: *dlloyd@trincc.bitnet*

The concept of representation is perhaps the single most useful idea in cognitive science, and Dennett & Kinsbourne (D & K) apply it well in their diagnosis of anomalies of subjective experience of sequence. Representations can misrepresent; vehicles of representation need not share properties with what they represent (specifically, representations need not mimic the spatial and temporal properties of their objects). Either of these ideas is sufficient to undermine the extravagant conclusions (of Libet and others) criticized in the target article.

Descartes, pioneer of the good idea of the representational mind, also left us several bad ideas. D & K indict the idea of a Cartesian bottleneck, a single central theater in which the brain projects its information into subjective consciousness. Two Cartesian ideas are combined in the theater metaphor: The singularity of states of consciousness, one per brain per instant, and the introspective interiority of states of consciousness, each

a state of self-consciousness arising from and directed at another, separate, brain state. D & K effectively summarize the case against singularity, but are less explicit on introspection. For example, the authors state that the onsets of many distributed "localized discriminations" and "content-fixations" "do *not* mark the onset of awareness of their content. It is always an open question whether any particular content thus discriminated will eventually appear as an element in conscious experience." The motivation for this distinction may be a desire to undermine the correspondence between subjective (reported) sequences and the underlying sequence of sensory discriminations, introduced in the discussion of the Libet experiments. Nonetheless, the general skepticism about representation of judgments and the mistrust of verbal reports as an absolute index of consciousness suggest that D & K are no friends of consciousness-as-introspection.

Beyond these important critical contributions, do D & K advance a positive theory of consciousness? I think that, contrary to their claim, the multiple drafts model is not a theory of consciousness, for the reason suggested by the quotation above: Some parts of the multiple parallel streams are conscious, and some are not. In this unexplicated distinction of conscious and unconscious cognition, D & K run the risk of replacing the central Cartesian theater with the equivalent of a suburban cinema complex, wherein many parallel streams of information project onto various screens of conscious awareness. The bottleneck may be gone, but the problem of consciousness remains at each of cinemas I through N. Ironically, D & K's deconstructive strategy – emphasizing the conspiracy of the where and when of consciousness – can afflict each rough draft in the multiple drafts model as surely as it undermines the Cartesian theater. Where does a local discrimination begin and end? When does the rough draft of an experience form? The problem of spatial and temporal smearing persists all the way down. Consider an arbitrary neuron in the visual cortex. Should its activity be regarded as a single cognitive event or process, or part of an event involving other neurons? If the latter, should this neuron be grouped with others in its layer? Its column? Its inputs and outputs? Should its activity be regarded as part of a state of consciousness?

From the difficulty of these questions – a spatial and temporal smear campaign – D & K correctly conclude that the distinction between conscious and unconscious processing is arbitrary and "lapses at close quarters." At this point one might be tempted by some form of antirealism about consciousness, and indeed similar arguments recur in Dennett's work, undermining naive faith in the reality of a wide variety of cognitive states. Unfortunately for antirealism, all such arguments collapse in the face of the reality of conscious experience, as undeniable for us as it was for Descartes. There are states of conscious awareness. The job of a theory of consciousness is to say what consciousness is. The multiple drafts model may identify an important property of consciousness, but fragmenting consciousness neither dissolves nor explains it.

The positive theory, however, is at hand. D & K argue that no effective distinction can be drawn between conscious and unconscious cognition. Why not view this result positively? If there is no distinction between conscious and unconscious representation, why not just identify consciousness and representation? An identity theory of consciousness would hold that conscious processing is cognitive (representational) processing, that conscious experience and cognitive science offer different descriptions of one and the same complex process. This is a substantive view, offering a genuine account of the nature of consciousness, and it has many independent virtues (see Lloyd 1989; 1991). Dennett and Kinsbourne may balk at the proposal on the grounds that any complex cognition results from elaborate preprocessing, necessitating swarms of unconscious micro-representations. But their parsimonious rejection of unnecessary *re*-presentations should also undermine unnecessary *pre*-presentations. Models of distributed representation and processing, abundant in both connectionism and cognitive neuroscience, indicate the power of simultaneous soft constraints to mold complex recognitions with very little preprocessing of inputs. Narrative consciousness might well frame its first extemporaneous drafts on otherwise blank pages.

The target article undermines *every* attempt at a functional distinction between conscious and unconscious processing. I for one will be very interested to see what Dennett & Kinsbourne are willing to conclude from their own well-argued premises.

UnCartesian materialism and Lockean introspection

William G. Lycan

Department of Philosophy, University of North Carolina, Chapel Hill, NC 27599-3125

Dennett & Kinsbourne (D & K) have made a solid case against Cartesian materialism. In particular, I am persuaded that relative to present evidence, the "turnstile of consciousness" assumption is unmotivated if not gratuitous, and that the famous "temporal anomalies" give us excellent reason to reject it in favor of the Multiple Drafts model.[1] My purpose in this commentary is only to assess the impact of the Multiple Drafts model on my favorite theory of consciousness, that is, Locke's introspective or "inner sense" account, as physicalistically implemented by Armstrong (1968; see also Lycan 1987, Ch. 6; 1990).

According to the inner sense theory, conscious awareness is the successful operation of an internal scanner or monitor that outputs second-order representations of first-order psychological states. But an "internal scanner" sounds awfully like an internal *audience* seated in a Cartesian Theater, even if it and the Theater are made of physical stuff; is the inner sense view not then committed to Cartesian materialism? And do D & K's arguments not bear negatively on the probity of introspection itself?[2]

It should be clear that the inner sense view itself is in no way committed to Cartesian materialism. For even if an internal scanner resembles an internal audience in some ways, the "audience" need not be seated in a Cartesian Theater: There need be no *single*, executive scanner, and no one scanner or monitor need view the entire array of first-order mental states accessible to consciousness. Accordingly, there need be neither a "turnstile of consciousness" nor one central inner stage on which the contents of consciousness are displayed in one fixed temporal order. An internal monitor is an attention mechanism that can presumably be directed upon representational subsystems and their stages; no doubt internal monitors work selectively and piecemeal and their operations depend on control windows and other elements of conative context, just as is predicted by the Multiple Drafts model. On this point, the inner sense theory has already parted with Cartesian materialism, and in the direction of D & K's view.

Indeed, the Multiple Drafts model *supports* the inner sense theory in an important way: It predicts introspective fallibility of two characteristic sorts. First, D & K emphasize that the result of a probe is a *judgment* made by the subject, a judgment that does not (or not eo ipso) simply report a "presentation" to an inner audience. And the "temporal anomalies" alone should have made us question the reliability of introspective reports. Introspection (Orwellian, Stalinist, or neuter) gets small temporal details wrong. That tends to confirm rather than to impugn the inner sense view of consciousness. If conscious awareness is indeed a matter of introspective attention and if introspection is the operation of a monitor or self-scanner, then such anomalies

were to be expected – for monitors and scanners are characteristically fallible on details, and D & K show admirably how such devices might corporately mix up temporal sequence in particular.

Second, the Multiple Drafts model predicts the nonexistence of a single "optimal time of probing," and (more drastically) ensures that probing "changes the task," that is, interferes with the very process it purports to be monitoring. That too is good news for the inner sense theory. For if introspection is the operation of a monitor or self-scanner, then revisionary effects of the present sort are again just what we should have expected; monitoring instruments (such as ammeters) typically do affect the values of the magnitudes they measure.

What, then, of introspection's own efficacy? Note carefully that this question is entirely independent of the inner sense issue. But note also that the scope of unreliability exhibited by the anomalies and the Multiple Drafts model is very small, tied to temporal differences within the tiny intervals involved, a small fraction of a second in each case. Thus, so far as has been shown, the fallibility is insignificant for most scientific and philosophical purposes even though it is crucial in sorting out the "temporal anomalies."

In any case, the inner sense theory of consciousness survives the collapse of Cartesian materialism, and is even strengthened by it.

NOTES

1. But for the record, I am completely *un*persuaded by D & K's claim that the Orwellian and Stalinesque accounts of (e.g.) color phi differ "only . . . verbal[ly]." Can't we finally call a halt to these creaking verificationist arguments? (Putnam 1965).

2. In an earlier draft of their paper which I was given in February of 1991, D & K made several arguments more specifically against the efficacy of introspection. Foolishly continuing to work with that MS instead of transferring my marginalia to *BBS*'s official preprint, I wrote a commentary criticizing the several arguments point for point. When I then checked my own commentary draft against the preprint for page-numbering purposes, I was aghast to find that each of the three sections on which I had focussed had been crisply excised – it seems temporally *before* I had written my commentary (!). At first this seemed magical, a precognition on D & K's part enabling them to avoid criticism by retroactive withdrawal; but then I realized it was only an instance of the Multiple Drafts phenomenon itself, indeed a better case than D & K's own word-processing example; more power to the Multiple Drafts model.

Little "me"

Drew McDermott

Computer Science Department, Yale University, New Haven, CT 06520
Electronic mail: *mcdermott@cs.yale.edu*

My problem with commenting on this target article is that I agree with it so completely and have trouble seeing how anyone, in this day and age, could disagree. I suppose one could take issue with the details of the Multiple Drafts model, but the details aren't really the issue. We find the world divided between those, like Dennett & Kinsbourne (D & K), who think that the computing brain is primary and the conscious mind derivative, and their traditionalist opponents, who think the opposite. For the first camp, the anomalies discussed by D & K are not really anomalies. There is just not that much to explain. It is too bad that they seem so bizarre to those who think that a primary fact about the conscious mind is that it is an infallible knower of itself. Unfortunately, those who persist in this quaint traditional attitude are in the majority.

I take the Multiple Drafts model, in essence, to be this: I am a character in a story my brain is making up. Consciousness is a property I have by virtue of my brain's attributing it to me. My story doesn't have to cohere completely to be useful. Sometimes it gets a bit ragged around the edges, as in the experiments discussed by D & K.

It's not hard to find other examples of the breakdown of the concept of the self. Consider the following true story: I recently had to badger some colleagues about some bureaucratic forms. I got a bundle of these forms, and went looking for the colleagues. I didn't find many of them, so I switched to a different errand. It involved carrying a very similar bundle of papers to a completely different destination. Along the way, I encountered one of the colleagues I had been seeking before. I told him I had been looking for him to give him a form, and proceeded to hand him one of the papers in my hand. I realized my mistake halfway through the gesture. I took the paper back, and went on my way. Then I met another wayward colleague, and went through roughly the same encounter again, including the mistake.

Question: On each occasion, at the moment I grasped the paper with the intention of unloading it, did I or did I not believe that I had grasped a copy of Bureaucratic Form A? To which the correct answer is: The concept of "I" has broken down in this situation. You can investigate the phenomenology of such an experience ad infinitum (Did I believe for an instant? Did I believe one way and act another?) but you'd be on a wild-goose chase. The proper explanation lies in tracing the flow of signals in my brain. Furthermore, the phenomenology is just another consequence of that flow.

The support for computational models of consciousness like Multiple Drafts is that they are the only current contenders for a materialist theory of consciousness. Against that support we have (by my count) three sorts of objection: introspective, epistemological, and ethical. The *introspective* objection is that I feel quite different from what the theory suggests I am. The *epistemological* objection is that all my knowledge is grounded in what I perceive, so "I" can't be derivative. The *ethical* objection is that if selves turn out to be constructs of information-processing systems, it will be hard to say why they are of any value.

These are serious objections, and one might be justified in heeding them, given the weak support for computationalist theories. But let me try to explain why the objections have so little force for me.

The introspective objection goes like this: I wake up in the morning. My eye opens, and a "point of view" pops into being. I ponder whether to try to go back to sleep or get up. Or do I only seem to ponder? Can it really be that there are several equally complex agents governing my behavior at such a moment, that conspire to invent this single person who is now thinking about getting up? Could they be inventing other persons at the same time, that I am unaware of? How many selves could my head hold?

Or put it another way: Suppose I met a person who behaved just like a real person, but was actually an unconscious zombie. Perhaps that's not possible, but if is, then presumably there would be a difference in the way that person's brain processed information. Upon inspection, one could find the exact place where *my* brain maintained a model of a conscious self, and *his* did not. But intuitively it seems that one could stare at my "circuit diagram" all day and still not believe that *that* was all there was to being conscious. Surely, you would speculate, one could have all that circuitry and still be a zombie (or still not have qualia, or whatever).

This is where you have to decide whether you are really a materialist or, in your heart, a dualist. Presumably, on any materialist theory, there is something like a circuit diagram (or a chemical reaction, or something) which goes through certain motions that constitute consciousness. Introspection is always going to be at odds with such a theory. The nice thing about a Multiple-Drafts-style model is that it explains exactly why the conflict seems so irreconcilable: The brain makes up a story in

which the main character is quite different from a piece of circuitry.

Even if you dislike computationalist theories, I don't see how you can take introspective evidence seriously and still be a materialist. Introspection long ago ceased to be an important tool of psychological experimentation because it became so clear how unreliable it is. Episodes like my adventure with the bureaucratic forms routinely expose cracks in our attempts to trust the introspective model of the self. The experimental results discussed by D & K drive wedges deeply into those cracks.

The second objection, the epistemological one, has the least force for me, but, I suspect, the most for the average philosopher. For hundreds of years, the problem of extrapolating from sensory experience to knowledge has been central to philosophy. A materialist theory of experience will make it necessary to rethink the epistemological issues *without starting from experience*. The basic data of knowledge were supposed to be facts like "I am experiencing a barn-shaped visual percept," from which one was supposed to infer conjectures like "There's a barn in front of me." If we take a computationalist model seriously, then this whole story is a fantasy about a fantasy. What actually happens is that (a) people are quite good at recognizing barns – i.e., they have reliable barn detectors; (b) they also, when required, automatically make up stories about how they do it, involving barn-shaped percepts.

So what's the problem? The epistemologist wants to ask how the person can ever be *sure* there's a barn there. He can't be sure, we answer, but if we perform experiments we discover he's right 99% of the time. No, no, that's not the problem, the epistemologist explains. The problem is, How can *I* be sure? I can't step outside myself and do the experiments on my brain and the alleged barns.[1]

This is where a computationalist theory of consciousness demands that you rethink the problem. The epistemological problem is not about how *I* can be sure, because "I" am not that important. I'm just something my brain made up. It seems to my self that it could exist without my brain existing, or any other physical object, but this impression turns out to be an error. Once it's explained what selves actually *are*, it will be clear that they require a computational substrate to exist. Solipsism as even a logical possibility will fade away. A self can no more exist by itself than a wave can exist without water.[2]

The dethroning of the self brings us to the third objection, ethics, which is the most crucial. Perhaps the consequences of believing D & K's theory are so awful that even if it is true we should suppress it. It has always seemed axiomatic that if I care about nothing else in the world, I can about "the dear self."[3] If the self is just a construct, then who cares whether it cares about itself? But what else is there to care about?

I don't know how to answer this objection, but we had better think of an answer, lest the twenty-first century be even more wasteful of human life than the twentieth. We used to be able to say things like "I impose cognitive categories on the world (space and time), and I impose ethical categories, too." But it is a consequence of materialism that such powers are far beyond little "I." If people are valuable, it is not because they are imperishable souls connected to bodies only for a brief sojourn. They have to be valuable for some other reason. For now, we just have to take it as a postulate that creatures that invent conscious selves are to be cherished and protected more than other information-processing systems. We're not going to get very far with inanities like "All life wants to survive and reproduce," which, even if were true, would only explain why people keep going when they don't stop to think. If we don't find a better ethical foundation, then we'll have to conclude that the examined life is not worth living.

Kant used to claim that he had wrought a "Copernican revolution" in philosophy, but it is becoming clear that what he actually achieved was a "Ptolemaic counter-reformation" that postponed the true revolution for a couple of centuries. Now a real Copernican shift is upon us. Such things are never pleasant.

NOTES

1. I realize that few philosophers nowadays would accept such a naive statement of the classic empiricist puzzle. So substitute a more sophisticated version, such as: How can *I* achieve intentionality – the ability to think about a particular object outside myself – when a *brain* can only process signals and symbols?
2. Thanks to Eric Mjolsness for pointing out this analogy.
3. Kant, *The Critique of Practical Reason.*

Conscious versus unconscious processes: Are they qualitatively different?

Eyal M. Reingold

Department of Psychology, University of Toronto at Erindale, Mississauga, Ontario, Canada L5L 1C6

Electronic mail: *reingold@psych.toronto.edu*

Can conscious and unconscious processes be distinguished empirically? No amount of precision in measurement is in itself sufficient. Equal emphasis on methodological rigor and conceptual clarity is necessary for genuine progress. Methodologically, the study of the unconscious involves illusive and fragile phenomena, difficult to replicate findings, and inconsistent and sometimes inadequate measurement criteria (see Eriksen 1960; Holender 1986; Merikle 1982; Reingold & Merikle 1988; 1990). Conceptually, definitional and terminological chaos and implicit assumptions that are rarely acknowledged fuel the often futile controversy between "believers" and "nonbelievers" in the unconscious (see Dunn & Kirsner 1988; 1989; Erdelyi 1985; 1986; Lockhart 1989; Reingold & Merikle 1988; 1990; Richardson-Klavehn & Bjork 1988). Not surprisingly, these difficulties have prompted practical researchers to stay clear of the conscious/unconscious distinction (Searle 1991), while related phenomena have been investigated under a variety of different dichotomies (e.g., attended/unattended, intentional/incidental, explicit/implicit, overt/covert, controlled/automatic, etc.). Yet the terms consciousness and awareness still seem to preserve best the historical link among related ideas; moreover their very ambiguity is an accurate reflection and a constant reminder of the vagueness of the theoretical constructs and their relation to empirical work in this area.

It is precisely in this context that Dennett & Kinsbourne (D & K) make a truly unique contribution. They not only expose brilliantly an important confusion between the temporal properties of the process of representing and the temporal content of the representations themselves, but they also provide powerful metaphors that may help one avoid sliding back into this ingrained confusion.

An inadvertent consequence, however, of D & K's very compelling case against attempts to time the emergence of consciousness and against the Cartesian notion of a transition between the unconscious and the conscious realms, may be an underestimation of the importance of the conscious/unconscious distinction. Although one can agree with the authors' portrayal of the graceful interplay between conscious and unconscious processes, and its temporal indeterminacy, this should not prevent us from trying to explore and explain the qualitative differences between these processes. This line of work would complement D & K's arguments; their very effective critique of nonquestions should be coupled with alternative questions. It is precisely because consciousness cannot be temporally or spatially localized in the brain that identifying qualitative differences between conscious and unconscious processes is crucial for the conscious/unconscious distinction. In-

deed, the method of establishing such differences (e.g., Cheesman & Merikle 1986; Dixon 1971; 1981; Jacoby & Whitehouse 1989; Jacoby et al. 1989; Merikle & Reingold 1990; 1991; Shevrin & Dickman 1980) may provide the ultimate test for the value of the conscious/unconscious distinction. More specifically, if the difference between conscious and unconscious processing is quantitative rather than qualitative, then the merit of agonizing over this distinction becomes highly questionable (see Cheesman & Merikle 1986). Perhaps in their response, D & K could specify where they stand on this issue.

On a more general level, Libet's work (see Libet 1985a; 1987; 1989a), reconsidered by D & K, is an example of the dissociation paradigm for the study of the unconscious (see Erdelyi 1985; 1986). Briefly, this requires two indices of processing, one which reflects information available to consciousness and another which reflects information available irrespective of consciousness. As elaborated elsewhere (Reingold & Merikle 1988; 1990), when consciousness is defined in terms of subjective report or claimed awareness, as in Libet's research, the dissociation paradigm yields ample evidence for unconscious processing. However, there is no general consensus as to the adequacy of such a definition. In fact, no operational definition of conscious awareness can be justified solely on an a priori basis; each requires convergent empirical evidence to establish its validity. Thus, the qualitative-differences approach may help validate any proposed index of consciousness.

Content and conformation: Isomorphism in the neural sway

Mark Rollins

Department of Philosophy, Washington University, St. Louis, MO 63130
Electronic mail: *c34801mr@wuvmd.bitnet*

Dennett & Kinsbourne (D & K) make a compelling case against two assumptions: (a) There is a canonical description of the content of consciousness. (b) Temporal content must be represented by temporal properties of neural events. However, the argument leaves open the possibility of a role for temporal isomorphism. D & K argue primarily against the general need for isomorphism and secondarily against the plausibility of it. The primary attack is convincing; but two points have to be reckoned with for the second: (1) The necessity claim can be seen as inherently task-relative. Isomorphism may be required for some temporal tasks only, not across the board. (2) The appeal to isomorphism need not commit the canonical fallacy. Thus, isomorphism cannot be seen as incredible merely because it keeps bad company.

1. Indeterminacy of content goes all the way in; there are many accounts that the brain itself might construct that are equally compatible with the data available to it. Of course, we do behave as if our experience made sense to us. Though multiple representations of conscious experience can coexist, they could not, without pathology, all actually constitute consciousness simultaneously. Thus, anticanonical arguments notwithstanding, D & K can claim that there is a "natural" interpretation for the brain to settle on; namely, a state that is typically the proximal cause of the subjects' reports. The state is natural because it is an inference caused to occur by the events in question, guided by the need to simplify and explain. But the fact that people impose a reportable interpretation does not mean it is the best possible one.

The temptation, then, is to try to rescue a canon by relativizing privileged versions of experience to passing "situations:" One version will rule, but it's one per situation. Unfortunately, the taxonomy of tasks that would distinguish different situations is underdetermined and indeterminate, too; no matter how narrowly a task is circumscribed, there will always be more than one good version of the experiences that occur in its performance.

But this raises an obvious problem. D & K give a reluctant nod to isomorphism by conceding that content is partly determined by when representing begins and ends. But, they argue, initiating constraints only matter early in the representational process (they can be disregarded later); and temporal control boundaries are only significant for certain kinds of task.

The problem is that, given the indeterminacy of tasks and D & K's emphasis on changes from moment to moment, there will be no single beginning and ending points, no one temporal window on the brain's representing. That might be thought to make *any* concession to isomorphism unnecessary. But the point cuts both ways: If the need for isomorphism is taken to be task-relative, then a defender can argue for an extensive role for it by arguing independently for the right sort of task parameters: narrow subevents, with significant beginning and ending constraints, on which larger events depend.

2. It's important that no best isomorphism need be implied; no "optimorphism," as we might call it. The unspoken assumption is that order is the one sort of temporal property that matters; and since there are objective constraints on it, any isomorphism between order and representation must be singular. The apparent counterargument is that since representation of order is not singular isomorphism should play no role in conscious experience.

However, there are at least *two* types of task in these experiments, event identification and temporal location. They are analogous to object identification and spatial location. These tasks are not reducible to one another. Events like tappings and light flashes have duration and thus temporal identities and they belong to an ordered set. But the duration of an event need not be established by fixing exactly a sequence of other events, nor do gross before-and-after relations require precise event identities. This makes possible a variety of temporal "perspectives." Faced with this, the brain can be seen as a strategy user whose strategy requires attaching authority to its own favored interpretations; a provisional *stipulation* of authority for the guidance of behavior.

In that case, the brain might authorize interpretations via isomorphism, if their authority is somehow secured thereby. There is at least one mechanism that could account for this tendency: The brain may be vulnerable to encoding effects. Specific differences in ability can derive from the way information is processed during learning. For example, word recognition is enhanced for action verbs if the action is mimed, for nouns if the denoted object is observed (Englekamp 1988). The effects are item-specific. And the differences do not require overt behavior or actual observation; internal simulations can produce them as well.

In some cases, then, the items might be events and movement sequences. Learning is not obviously involved here; but something like encoding effects could occur during the selective evolution of neural capacities. Of course, nature could economize with an amodal system. But nature tempers frugality with provision for individual difference, a known aid to evolution. Mode-specific encoding effects might result merely from the use of precise configurations as opposed to gross generic ones, all with a common interpretive "vocabulary." It isn't clear, however, why we should think that is the case, except to preserve amodalism in a way that begs the question. Moreover, this seems to miss the spirit of nature's interest in real diversity.

If the motor encoding of actions translates into different sorts of neural representings compared to the sensory encoding of perceived objects, then the experience of temporal order in movement sequences can translate into forms of representing different from those of seeing events. As D & K suggest in one analogy, the inconsistency of a loud request for quiet may be

effective by calling attention to the authority and power of the speaker. But an exaggerated whisper can amplify the force of the same content, too. The need for complementary isomorphism is not for the sake of representing per se; it is for an effective strategy with limited means. In any case, isomorphism need not work alone or have the final say. Given our repertoire of resources, it may be part of nature's best effort so far, a compromise solution.

Time and consciousness

David M. Rosenthal

Ph.D. Program in Philosophy, City University of New York, Graduate School, 33 West 42 Street, New York, NY 10036-8099
Electronic mail: *drogc@cunyvm.bitnet*

What is it for a mental representation to be conscious? It is a familiar Cartesian doctrine that being conscious is part of what it is for a state or a representation to be mental. Since being conscious is part of being mental, not only are all mental states conscious; being conscious must be an intrinsic property of mental states. A state's being intrinsically conscious captures the idea that, as Descartes put it, we are immediately conscious of all mental states; nothing mediates between a mental state and our being conscious of it.

Given these Cartesian assumptions, there is no room for the brain (or mind) to reinterpret or revise the way I am conscious of my mental states, since my being conscious of them is intrinsic to the states themselves. Once a mental state exists, its very nature fixes what my being conscious of that state can tell me about it. Because being conscious of mental states is immediate and is intrinsic to the states, what it tells us about our mental states is not only the final draft on that subject, but the only draft.

All this has consequences for subjective temporal succession. Suppose I first see event *a* and then event *b*. If being conscious is intrinsic to every mental state, both cases of seeing will, at the exact moment they occur, be conscious states; simultaneously with seeing each event I will be conscious that I do. I will be conscious of seeing *a*, and then conscious of seeing *b*.

Consciously perceiving temporal succession between *a* and *b* is part of the way my seeing *a* and seeing *b* are conscious perceptions. So if being conscious is intrinsic to those perceptions, there's no room for the mind (or brain) to adjust how I consciously perceive that temporal succession. Put another way, if a state is intrinsically conscious, then the state represents its own occurrence; so there can be no disparity between the timing of a state and that timing as the state represents it. There will be no alternative, then, to my seeing *a* as having occurred before *b*. Mellor (cited in D & K sect. 2.1, para. 11) will be right that we perceive the subjective temporal succession of two events simply by way of the objective temporal succession of our representations of those events.[1]

If mental states were all conscious, being conscious might well be an intrinsic property of those states. But it's widely recognized that many mental states are not conscious. That they are not would be hard to explain if being conscious were an intrinsic property of mental states. So we must reject the idea that being conscious is an intrinsic property of mental states.[2]

The one thing that's uncontroversial about a mental state's being conscious is that it involves one's being conscious *of* that state in some way or other. Let us call our being conscious of our conscious states in the relevant way *transitive consciousness*, since it's a case of being conscious *of* something. If being conscious is an extrinsic property of mental states, our being transitively conscious of those states will also be extrinsic to them. A mental state will be distinct from our being transitively conscious of it.

The idea that mental states are distinct from our transitive consciousness of them fits well with D & K's Multiple Drafts model. When we are conscious of something, we are conscious of it under certain aspects and not others; we represent the thing we are conscious of in certain ways, and not others. So it is with our being conscious of our mental states. Being transitively conscious of a mental state means representing that state in a certain way, and how we represent it will determine what sort of state we think we're in. [See also Searle: "Consciousness, Explanatory Inversion, and Cognitive Science" *BBS* 13(4) 1990.]

How we represent the things we are conscious of, moreover, can change over time, and there is no reason why this too should not happen with our transitive consciousness of our mental states. Our being transitively conscious of our mental states involves representing them in certain ways. These representations can change, and as they do, corresponding changes will occur in what mental states we seem to be in. In effect, we'll have a series of drafts about the contents of our minds. Since how transitive consciousness represents our mental states is distinct from those states, these changes need involve no shift in the nature of the mental states themselves; all that has to change is the character that our transitive consciousness of those states represents them as having. Moreover, we will not consciously notice these changes, since it is only in virtue of how our transitive consciousness of our mental states represents them that we are conscious of those states at all. The latest draft will seem, for the purposes of consciousness, to be the only draft.

Things would be different if our transitive consciousness of our mental states were intrinsic to those states. How we are transitively conscious of them would then be part of their nature. Our transitive consciousness would occur simultaneously with the state, and the way our transitive consciousness represents the state could not change without a change in the very nature of the state itself. But apart from discredited doctrines about having infallible or exhaustive access to our mental states, we have no reason to think that our transitive consciousness of our mental states is intrinsic to them.

Let us turn again to subjective temporal succession. Consciously perceiving a temporal succession between *a* and *b* involves representing my perception of temporal succession in a particular way. Since my transitive consciousness of that perception is extrinsic to it, nothing about my perceptions of *a* and *b* fixes the way I represent my perception of their temporal relations. In particular, the objective temporal succession of my perceptions of *a* and *b* cannot fix how my transitive consciousness of my perception represents that perception of temporal succession.

Moreover, the way my transitive consciousness represents that perception can change over time.[3] The way my transitive consciousness of my perceptions represents them determines what perceptions I seem to have; so the latest transitive consciousness will seem to be the only one I've ever had. This is so even when I am perceiving temporal succession. So if the way I represent my perception of temporal succession changes, it will seem that that is the only way I ever perceived things. This is just what is needed to explain the cutaneous rabbit and the phi phenomenon.

According to the foregoing model, one's transitive consciousness of one's mental states is distinct from those states and can change independently of them. And it is this transitive consciousness of one's mental states that determines what one takes those states to be. Is this model Stalinesque or Orwellian?

For the Stalinesque theorist, perceiving something fixes my memories, but I can edit my perceptions; for the Orwellian, I can eradicate the effect of my initial perception by rewriting my memories. If by "perception" we mean conscious perception, the model I have put forth may look Stalinesque. A conscious perception is a perception plus the transitive consciousness of it; so the brain can edit my conscious perception by altering how

my transitive consciousness represents the perception. But given the latest version of my transitive consciousness, my conscious perception may well fix subsequent memories.

We might instead mean by "perception" just the nonconscious perceptual state, apart from any transitive consciousness of it. The foregoing model may then seem Orwellian. Without transitive consciousness, perceptions don't by themselves determine subsequent memories, since memories will follow the way my transitive consciousness of my perceptions represents them. The model I have proposed gives us no reason to think we can edit these nonconscious perceptual states. But if we can edit our transitive consciousness of our perceptions, surely we can alter our memories of them.

The deeper issue, however, is whether whatever revising does take place should count as an alteration of memories or of initial perceptions. Here the foregoing model resists easy classification. The brain edits our conscious perceptions by changing the way our transitive consciousness of those perceptions represents them. This can happen earlier or later. The Stalinesque theorist thinks we edit our perceptions, while the Orwellian maintains that we rewrite our memories. So the brain's revising of our transitive consciousness of our perceptions will seem more Stalinesque the earlier it happens, and more Orwellian the later it happens. The line between Stalinesque and Orwellian is, as D & K insist, arbitrary. Since the revising of our transitive consciousness can happen earlier or later, at small time scales we won't be able to draw a principled line between what's "not yet observed" and what's "already observed."

If being conscious were intrinsic to mental states, it would be natural to fix the time each mental state occurs by the time at which it becomes conscious. Can we do that with those mental states that are conscious, even though being conscious is not an intrinsic property of them? We could not reliably fix comparative timing this way, since there is no single place in the brain at which the transitive consciousness of every mental state occurs. Still, we might be able to draw a nonarbitrary distinction between rewriting memories and editing conscious perceptions.

This move is unavailable. Our transitive consciousness of a perception can be revised; so which transitive consciousness should count for purposes of timing? Is it the first, which fixes when the perception initially become conscious? Or the last, after which no more changes occur in the way our transitive consciousness of the perception represents it? Or should we pick the time, possibly in between first and last, at which the way our transitive consciousness represents the perception becomes relatively stable? Since there is no principled answer to these questions, we cannot time mental states by reference to the time of their being conscious.

Finally, are Dennett and Kinsbourne right that there is no "fact of the matter about exactly when (in 'absolute' time . . .) a conscious experience happens" (sect. 3.1, para. 12)? That depends on what we mean. We have no reason to doubt that we can fix the time of whatever mental state we're conscious of, independently of our transitive consciousness of it. But the brain can revise the way our transitive consciousness of any experience represents that experience, and each successive transitive consciousness is subjectively no less authoritative than the preceding ones. So if by "conscious experience" we mean the experience plus our transitive consciousness of it, there is indeed no fact of the matter about its timing.

NOTES

1. Mind-body dualism plays no role in this line of thinking; as Kant in effect showed, the felt need to postulate mental unity is independent of such dualism.

2. For more on why being conscious is not an intrinsic property, see Rosenthal (1990; 1991).

3. Perhaps, as Dennett & Kinsbourne suggest, in order to find the best temporal fit among the contents of the representations of those events (sect. 2.1, para. 8).

Cinema 1-2-Many of the Mind

Adina L. Roskies[a] and C. C. Wood[b]

[a]Salk Institute, MNL-O, La Jolla, CA 92037 and [b]Biophysics Group (P6), Los Alamos National Laboratory, Los Alamos, NM 87545

Electronic mail: *[a]adina@helmholtz.sdsc.edu and [b]wood@tailor.lanl.gov*

One gets the impression that Dennett and Kinsbourne (D & K) are busily nailing the last few tacks into a big *Out of Business* sign stretched across the Cartesian Theater. However, upon closer inspection the sign discloses a far less definitive message: *Closed for Remodeling – Reopening Soon as Cinema 1-2-Many of the Mind.*

The stated goal of D & K's target article is to expose and dismantle the Cartesian Theater. Although we agree with much of their diagnosis, in particular their central conclusion that the neural events subserving conscious experience are spatially and temporally distributed and that there is no single "where" and no single "when" to look for them in the brain, we have serious reservations about the Multiple Drafts model proposed as an alternative. The Multiple Drafts model not only fails to close the Cartesian Theater as D & K intend, it retains and multiplies some of the deficiencies of that model by replicating the major mysteries of conscious experience across an indefinite number of ill-characterized Multiple Drafts. In effect, D & K are replacing the single Cartesian stage with a multi-screen Cinema 1-2-Many of the Mind.

The major virtues of the Multiple Drafts model are what it says about what conscious experience is *not:* (a) that there is no single where and when of conscious experience in the brain and (b) the temporal properties of conscious experience need not correspond with the temporal properties of the neural events that mediate conscious experience. However, when it comes to proposing what conscious experience *is,* the Multiple Drafts model is decidedly silent about a number of key issues:

1. Why do some neural processes constitute "drafts" having content that can contribute to conscious experience while others do not?

2. How do the various layers of D & K's mental palimpsest interact, compete, and gain primacy to produce the sense of a (quasi-) coherent series of perceptions, intentions, and actions that characterize our interaction with the world?

3. What constitutes "editorial revision" and who/what does the revising?

4. How is the metaphor of a "draft" an improvement over the metaphor of a theater in dispelling the "infinite regress of too-powerful homunculi"?

D & K are, of course, not alone in having a less than satisfactory account of conscious experience and its relationship to the brain. The chasm between the subjective and the objective has stumped thinkers since Descartes and we confess to finding ourselves in exactly the same muddle. So what are would-be theater-goers to do for entertainment in the faced of D & K's "*Out of Business*" sign? Are we to concede that the "where and when of consciousness in the brain" are totally outside the bounds of science-as-usual? Not quite, for there are directions to proceed in which considerable progress, albeit somewhat indirect, can be made.

One promising direction to proceed is down. Although investigations of the brain do not address the subjective-objective relationship head-on, any attempt to characterize this relationship will benefit from a more thorough understanding of either the subjective or objective component taken separately. Neuroscience has only begun to scratch the surface of the deep mysteries of the brain and we are convinced that science-as-usual will reveal that many puzzling phenomena, including some of the temporal paradoxes that D & K view as particularly problematic, are understandable consequences of neural structure and function. Just as visual spatial illusions have been treated as discrepancies between the subjective and objective

that can be informative about the properties of the visual system, we believe that D & K's "temporal anomalies" should be viewed as illusions in the temporal domain that can be informative about the normal operation of the system (cf. Churchland 1981a), not as reasons to invoke a model like Multiple Drafts which only exacerbates the muddle. In the remainder of this commentary we outline three examples of phenomena at the level of neuroscience that may help shed light on the "paradoxes" investigated by D & K. Although these examples do not address the subjective–objective relationship directly, they can provide significant constraints on the nature of that relationship.[1]

Color-filling. The eye has a natural jitter that causes the visual field to "dance in place" on the retina. When images are artificially stabilized on the retina, imaged contours appear to fade and colors spread to fill spaces delineated by unstabilized boundaries. This phenomenon is called color-filling. The observer perceives colors that are not present in the corresponding location in the retinal image and fails to perceive colors that are retinally represented. Somewhere between the level of the retina and what is accessible to conscious experience a significant change occurs in the neural representation of color information. In their analysis of Orwellian versus Stalinesque "revision," D & K argue that the nature of the "revision" is forever inaccessible: We (either as observers or scientists) can never know whether the brain really "fills in" the perceived color, or whether it sends an executive signal indicating "color filled." To the extent that this question is posed relative to the assumption of a single time and place in the brain subserving conscious experience, we agree with D & K that it is unresolvable. But if we broaden the question to "Where and when in the brain does the representational change occur?" then there is constructive work to be done. Neuroscientific evidence suggests that the visual system actually constructs the filled-in information by the time visual cortex is reached (Piantanida 1985; Piantanida & Larimer 1989). Neurons in visual cortex whose receptive fields coincide with the location of the filled-in but not-actually-present-at-the-retina color demonstrate the same adaptation effects as neurons that actually see that color. Thus, the illusory color information is constructed and present in visual cortex as color information of the usual sort. Color-filling therefore provides an example of an identifiable modification of informational content that provides a candidate explanation for observers' reported subjective experience. One cannot, of course, conclude that these particular neuronal effects are the only basis or the most important one for the subjective experience of color-filling. As D & K emphasize, the neuronal events that contribute to the subjective effects may be distributed widely in multiple regions of the brain. Like many methods in neuroscience, single-cell recording techniques suffer from significant sampling limitations and the reported single-cell "color-filling" effects could be merely distant and indirect manifestations of the mechanisms that subserve the subjective effects. Nevertheless, what we do know from these results is that the brain has the capacity to generate color-filling at the neuronal level in visual cortex and this knowledge can place useful constraints on theories of color perception.

Apparent motion. Although we find it natural to think of cells as having spatial receptive fields, the notion of temporal receptive fields in vision is more rarely discussed. However, just as neurons have morphological properties that enable them to code for and respond to particular spatial patterns, their rich array of dynamical properties such as membrane capacitance and channel kinetics make them well-suited for computations in the time domain as well. Neurons in visual areas V1 and MT respond both to real movement (i.e., stimuli moving continuously across a neuron's receptive field) and apparent movement (i.e., stroboscopic presentation of stationary stimuli). Moreover, their responses vary systematically with spatial and temporal parameters of the stimuli that closely parallel the psychophysically defined limits of apparent motion in human observers (Newsome et al. 1986). Again, one should not overinterpret these results, either with respect to their necessary contribution to subjective effects or their representativeness of neuronal mechanisms of real or apparent motion in general (cf. Hildreth & Koch 1987). Nevertheless, these single cell results, together with knowledge of the temporal properties of neurons, can lead to interesting and testable hypotheses. For example, the perception of apparent motion might begin with an initially nondirectional priming from the first flashed stimulus. If another stimulus with similar characteristics should appear within the primed region before the activation has decayed, a motion signal might be generated, either by potentiating the intervening circuit or by sending a signal that is interpreted by other areas as coding motion. As D & K point out, color phi is a special case of apparent motion, and a mechanism similar to the one operative in color-filling may serve to create a spread of color information which parallels the spatial interpolation.

Simultaneous manipulation of neuronal and subjective effects. D & K are acutely aware of the difficulties attending attempts to link brain activity and conscious experience, but their treatment obscures the important point that, objectively speaking, there is *a fact of the matter* about brain states and processes, not multiple, indistinguishable versions of facts. In the color-filling and apparent motion examples discussed above, the experimental strategy is to attempt to find cells or cell populations that bear relationships to stimulus manipulations similar to those reported by human observers in psychophysical experiments. As we have noted, such correspondences can provide *candidate* explanations from which one can begin to form hypotheses about neural substrates of a given subjective effect. However, one can go further and test such hypotheses by experimentally manipulating the neuronal populations in question using stimulation, lesion, pharmacological, or other techniques. For example, in the motion domain, Newsome and colleagues (Salzman et al. 1990) have recently demonstrated that a monkey's behavioral responses in a motion judgment task can be influenced by local microstimulation in area MT, the cortical region in which neurons appear to be especially sensitive to real and apparent motion. We emphasize that the results of such manipulation experiments do not necessarily indicate that area MT is *the* or even *a* system that subserves subjective experience of motion. Nevertheless, by systematically applying this experimental strategy across a variety of visual areas it may be possible to assemble a catalogue of regions that do and do not appear capable of biasing subjective experience (as inferred from behavioral report) according to the experimental manipulation employed.

In summary, although we readily acknowledge that neuroscience is unlikely to provide a bridge across the subjective-objective chasm in any direct sense, we believe that it can provide a much richer body of data from which to formulate bridging hypotheses than the sterile, no-fact-of-the-matter position advocated by the Multiple Drafts model. Data and theories from neuroscience can place significant constraints on what the subjective-objective relationship might be and perhaps allow us to address several important questions sorely lacking treatment in the Multiple Drafts account:

1. Which neural processes (where and when) appear to covary with subjective report and which do not?
2. Which regions and processes can we manipulate to bias subjective report and which not?
3. What characteristics of these processes distinguish them from processes that cannot be biased in this manner (i.e., what counts as a "draft")?

We agree wholeheartedly with D & K that the Cartesian Theater is doomed to fail. But after remodeling by the Dennett & Kinsbourne Co., what we have instead is hardly more radical

than a series of minitheaters. In short, after (rightly) discarding the assumption that there is a single where and when of conscious experience in the brain, the Multiple Drafts model seems to incur the most serious remaining problems of the Cartesian Theater, as well as new ones of its own. Modernization of old standbys will not result in the conceptual breakthroughs necessary to understand consciousness and its relation to the brain. What is really lacking are new forms of entertainment that genuinely dispel, not just obscure, the insidious Cartesian homunculus. In the Multiple Drafts model he can be found sitting in any or all of the minitheaters, this time clothed in the dapper attire of the Editor(s) of the Drafts.

NOTE

1. Neuroscience also provides abundant evidence in support of D & K's point that "we must distinguish features of representings from the features of representeds" in the brain's representation of temporal information. For example: (a) the frequency of an acoustic stimulus appears to be represented in the nervous system both as spatial and temporal codes (e.g., Javel et al. 1988); (b) interaural time (phase) differences that convey information about sound localization appear to be computed by temporal comparisons at one level of the owl's auditory system and represented as a spatial map at a subsequent level (Konishi et al. 1988); and (c) complex acoustic features mediating biosonar, including Doppler sensitivity, appear to be organized into spatial maps in the mustached bat's auditory cortex (Suga 1988).

Mental representation: Always delayed but not always ephemeral

Roger N. Shepard

Department of Psychology, Stanford University, Stanford, CA 94305-2130
Electronic mail: *roger@psych.stanford.edu*

There is much that is right about Dennett & Kinsbourne's (D & K's) likening of mental representation to unending revision of "multiple drafts" concerning what is going on in the external world. After noting two points that are in particular agreement with ones I have previously advocated, I must, however, draw attention to two points of divergence.

Representation of physical properties versus physical properties of representation. Certainly, I agree (with the target article, sect. 2.1) that *the representation of timing* is not the same thing as *the timing of representation* and that *the experience of (A before B)* does not entail *(the experience of A) before (the experience of B)*. In statements about mental events generally, terms may not be commutative or distributive. As Kubovy (1983) succinctly observed in reviewing work that my associates and I had reported on mental rotation (Shepard & Cooper 1982), to *imagine (a rotation of (an object))* is not necessarily to *rotate (an image of (an object))*. The second statement in each of these three italicized pair entails the problematic attribution of physical relations (spatial orientations or temporal orders) to nonphysical objects (mental images or subjective events). In contrast, the first statement in each pair focuses on the representation of the physical properties of physical objects and, in so doing, provides both for the objective study of mental representations in the psychological laboratory (Shepard 1990) and for the natural selection of representational systems in the physical world (Shepard 1987).

The distributed nature of the physical processes underlying representation. Also in agreement with D & K, I once characterized the brain as "a community of pretuned monads that come into harmonious action, with each other and with the world outside, through many glasses darkly" (Shepard 1984, p. 438–439). Implicitly equating "darkly" with "slowly," I further noted that "the necessarily finite velocity of signal propagation within a body must limits its processing of information" (Shepard, 1984, pp. 426–27). Indeed, I have regarded this processing limitation as the source of the fundamental law of apparent motion (Korte's third law) according to which apparent motion requires slower rates of alternation for more widely separated stimuli (Shepard 1981; 1984; 1989).

The representational system, despite its inherent time delays, has been evolutionarily shaped to represent biologically significant events in the external world. As D & K observe, instructions to report the temporal order of the internal representational events themselves call, in contrast, for "judgment of temporality of a sort that does not normally play a role in behavior control" (target article, sect. 3.2). Not surprisingly, subjects' attempts to comply with such instructions can yield anomalous results when the temporal intervals between the relevant neurophysiological events are less than the propagation and relaxation times of the neuronal system as a whole.

The sometimes perceptually concrete character of retrospective representations. I do question, however, D & K's insistence that "drafts" concerning external events are always like abstract descriptions of "narratives" (target article, sect. 1.1); that "the brain doesn't actually have to go to the trouble of 'filling in' anything," perceptually; and that to do so "would be a waste of time and (shall we say?) *paint*" (target article, sect. 2.2). My coworkers and I have demonstrated that processes of mental rotation (Cooper 1976; Cooper & Shepard 1973) and apparent motion (Robins & Shepard 1977) are *analog* processes in the sense that they successively activate (or "fill in") intermediate states corresponding, in concrete detail, to what would be successive intermediate states in the external world (see Shepard & Cooper 1982).

In the experiment by Robins and Shepard (1977), for example, apparent motion was ambiguously induced by alternately displaying a vertical and a horizontal bar. Subjects reported seeing a single bar rotating 90° back and forth either through the upper right and lower left quadrants or, instead, through the upper left and lower right quadrants. Was the difference between these two distinct perceptions of the same external stimulus merely a difference in after-the-fact description? Pursuing this question, we briefly flashed a small probe dot at a randomly selected time and location along the circular path of experienced motion and asked subjects to indicate whether the dot flashed before or after the bar appeared to pass through the location of the dot. For dots that were presented in a quadrant through which motion was experienced, dots further along the path of that motion could be flashed after longer delays and still elicit the response "before" – just as if a bar had actually moved through that quadrant.

Even more direct, neurophysiological evidence for the analog nature of a type of mental rotation (cf. Shepard & Metzler 1971) has subsequently been reported by Georgopoulos et al. (1989). They found that the cortical firing patterns recorded from monkeys who were preparing for a rotational movement passed through patterns corresponding to intermediate orientations. If the brain does sometimes "go to the trouble" of such a concrete "filling in," there must be some adaptive value in representing significant external events in the most accurate, complete, perceptual and, as I shall now argue, memorable form.

Adaptive functions of enduring retrospective representations. Here, I question D & K's supposition that each "draft," reflecting "the situation at the time it is generated," is no more "correct than another" and, by implication, is never destined to attain a more enduring (metaphorically, "published") form (sect. 1.1). True, our evolutionary fitness depends on our ability to respond to challenging events before our representational system has had time to relax into a final interpretive state. Typically, however, events are followed by more quiescent interludes during which the evolutionarily shaped interpretational process may run its course. The resulting more stable representation, even if too late to mediate reactions to the immediately pre-

ceding events (Libet 1985b; Libet et al. 1979), may be more veridical and, hence, more worthy of preservation for the guidance of future behavior.

Certainly, natural selection must have favored accuracy in learning, memory, and generalization as well as in perception. For example, experienced motion, though illusory in the contrived situation of apparent motion, nevertheless represents the external event that is both most consistent with the retinal stimulation and most probable in a world in which material objects are conserved (Shepard 1984). Normally, the representation of such a motion yields not only a more veridical percept but also a more useful memory. In short, retrospective "drafts" that achieve a sufficient degree of concrete consolidation, answering the call of consciousness to "publish," may not soon "perish."

ACKNOWLEDGMENT

Research supported by the National Science Foundation grant BNS 90-21648.

In defense of the pineal gland

Robert Teghtsoonian

Department of Psychology, Smith College, Northampton, MA 01063
Electronic mail: *rtex@smith.bitnet*

The straw man stands near the top of the list of useful strategies for the introduction of new ideas. Whatever the objective merits of a new theory, it can be made to seem even more compelling by contrasting it with an alternative so flawed, or even openly ludicrous, that the reader must either endorse the author's proposal or support the patently absurd. I write this comment not to find fault with the Multiple Drafts (MD) model (which I in fact believe to be ingenious and useful) but to suggest that Dennett & Kinsbourne (D & K) might have been more charitable in their treatment of Descartes, and to argue that there is some evidence to support a modern version of his basic idea, a sort of conceptual pineal gland.

Of course Descartes has been scolded for his metaphysics for centuries and D & K have nothing new to add here. But in this context the objection to a distinct res cogitans is gratuitous. Whether discriminations are distributed and multiple, as in the MD model, or central and single, as in the Cartesian Theater, the option of positing an immaterial agency to process that discrimination still exists. The decision against exercising that option seems more a matter of metaphysical prudence than of some intrinsic merit in the model. I don't see anything in the MD model precluding the assumption of a million small homunculi, each responding to peripherally located discriminations. Descartes might argue that, in closing his Theater, D & K have, without admitting it, reserved the right to operate in the same building a large number of rehearsal rooms where assistant directors examine the talent and revise the material, communicating with one another by e-mail. More complicated? Yes. Metaphysically superior? I doubt it.

The nub of the issue has not so much to do with metaphysics as it does with the question of whether there are places in the brain that receive inputs from various sources and perform some coordinating transformation of that data. I think the important contribution made by Descartes is the recognition that at some level there must be mechanisms that serve such coordinating functions. Had he had the opportunity to know modern neuroanatomy, we may suppose that Descartes would have insisted on neither a single mechanism nor a single location for the collation and processing of incoming data. But whether the Theater is large or small, and whether there is one such locus or many, we need to know first whether there are any coordinating operations performed on inputs from diverse sources and, if the answer is yes, what the nature of these transformations is and what subsequent use is made of the results.

One line of evidence has been developed from the study of perceived magnitude, the line of research begun in the 1950's by the Harvard psychologist S. S. Stevens (1975), who asked observers to attend to simple signals varying only in intensity, and to make judgments of their perceived magnitudes. Whether the signals were pure tones of a given frequency varying only in amplitude, discs of white light varying only in luminosity, or weights varying only in mass (to give but a few examples), and whether observers made their judgments by assigning numerical values or by adjusting the intensity of some other stimulus dimension, the result was consistently found to be a power function relating the observer's judgments to the intensities of the stimuli presented. Of equal importance was the discovery by Stevens that each perceptual continuum (loudness, brightness, heaviness, and the like) has a characteristic exponent for the power function, taking values in the range from about 0.3 to about 3.

In various reports (Teghtsoonian 1971; 1974) I have noted that low exponents are typical for continua with large dynamic ranges, while high exponents characterize continua with short dynamic ranges, and that there is reason to believe that at some level these varying dynamic ranges are, through processes of compression or expansion as needed, mapped into a single scale of subjective magnitudes. Indeed, it is as if intensity information from all inputs is routed to a single monitor whose output defines the experience of subjective magnitude (how loud a sound seems to a listener and how bright a source of light appears to be), and it is the maximum range of outputs of this device that makes the brightest light subjectively equivalent to the loudest sound. It is the apparent subjective equality of the many different dynamic ranges that suggests the likelihood of a central monitor whose properties define the limits of perceived magnitude. (Another, more empirical way of stating the argument is that there is no value on any given continuum that cannot be subjectively matched by some value within the dynamic range for any other perceptual continuum. During the several years since this rule of complete and exhaustive intermappability was first proposed, there has been [to my knowledge] no disconfirming evidence.) If the idea of a central monitor for intensity information proves to be valid, it could qualify as a kind of Cartesian minitheater, and, could be placed in the pineal gland, the big toe, or anywhere else that may be implicated by neuroanatomical research.

I might add one further line of argument in favor of such a monitoring device, although this one has been the subject of criticism and claims of disconfirming evidence. In earlier publications I have noted that the several perceptual continua are distinguished not only by characteristic power law exponents, but also by marked differences in resolving power. Despite the inaccuracies embodied in the traditional version of Weber's Law, it is possible to specify an approximate percentage change in intensity below which no change is noticed, and this value, like the exponent of the power law, depends on the perceptual continuum being tested. Thus, human observers may be aware of relative changes in the intensity of electric shock delivered to the fingertips that would be quite undetected if expressed as changes in sound intensity. I have noted (in the sources cited above) that resolving power appears to be inversely related to dynamic range, with differential sensitivity being greatest for continua with the shortest dynamic range. And if differential sensitivity is ultimately defined by the resolving power of the central intensity monitor described above, and if just-noticeably changed stimulus ratios are subjectively equivalent across all perceptual continua, the inverse relation between DR and resolving power would be seen as further evidence for the existence of a single monitoring device for intensity.

Presumably similar monitoring systems exist for other at-

tributes, and the ways in which the outputs of these devices are assembled remain a matter of speculation. Indeed, as Dennett & Kinsbourne themselves observe, it is essential "to learn exactly when and where various informational streams converge." The hard empirical work will continue to focus on attempts to discover how each variety of information is processed, and the ways in which the results of that processing are assembled and combined. It seems quite plausible to me that, in pursuit of that goal, some variety of Cartesian mini-theaters may be needed.

Does the perception of temporal sequence throw light on consciousness?

Michel Treisman

Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, England

Electronic mail: *treisman@psy.oxford.ac.uk*

We are born naive realists, believing that we see the world as we do because that is how it is. But sooner or later we have experiences which may lead us to distinguish between the real world that physicists' descriptions refer to and a realm of phenomenological experience. We learn that sticks that "look bent" in water are not bent in the physicist's space. To a major extent, it is to the phenomenological aspect of perception that terms such as "conscious awareness" refer. The distinction between phenomenological experience (the content of "conscious awareness") and the world inferred by the sciences has a strong intuitive force, although it is understood in different ways by different people.

What is the relation between these two realms of description? Physical instruments could measure the mass, reflectance, and location of the word-processor before me and thus confirm its presence. In practice I am satisfied to note its beige space-filling bulk, red label, and low-pitched murmur. I have also been persuaded that intermediate between these two realms of description, and necessary for their concurrence, lies a special part of the physical world, a network of active neural mechanisms constantly engaged in processing the information coming in from the sense organs. The observer who has graduated from naive realism and discovered the role of neural mechanisms may ask: What is the relation between the elements of phenomenological experience and the corresponding neural events? We must admit that the best answer currently on offer is that this relation – let us call it Relation$_u$ – is undefined and that there is nothing that we know about it with certainty. If we suppose that this relation is a reliable correlation between two realms, we have dualism. If Relation$_u$ represents causal interchanges in both directions this may be recognized as interactionist dualism. If we define Relation$_u$ as "is an aspect of" (whatever "aspect" may mean) or "is a property of," or "is identical to," we have materialism. Dennett & Kinsbourne's (D & K's) target article attempts to assess what implications certain observations on subjective timing may have for our understanding of Relation$_u$.

D & K's target article stresses important points which we do well to be reminded of. The time represented by an element of phenomenological experience is not the time at which that element is generated – as the naive realist with respect to time would suppose – but the time to which it refers, just as beige is not the color of a conscious sensation but of my word-processor. (I leave aside problems about "representation" and "referring.") We do not need a central observer to reperceive the products of perception.

D & K's main arguments relate to four experimental results which have been supposed to raise problems in relation to the timing of conscious experiences or timing in conscious experience. They involve apparently paradoxical relations between subjective timing and physical events.

Timing is an old topic in psychology. Before discussing D & K's treatment of these results, it will be useful to remind ourselves of some relevant earlier observations. James (1890) recorded that it had long been noticed when "expectant attention" was concentrated on one of two sensations, the subjective experience of the other might be delayed. "Thus, to use the stock example of the books, the surgeon would sometimes see the blood flow from the arm of the patient . . . *before* he saw the instrument penetrate the skin. . . . There is thus a certain difficulty in perceiving the exact *date* of two impressions when they do not interest our attention equally" (p. 409). This effect was extensively investigated by psychologists as the phenomenon of "prior entry." If a subject watched a pointer move over a dial and at some time in its traverse a click occurred, or if a click and a brief tactual stimulus were both given with the interval between them varying, and in each case the subject was required to direct his attention to one stimulus of the two, then that attended stimulus would be judged to occur earlier in time relative to the other than the physical difference in their times of presentation warranted (Boring 1950; James 1890; Stone 1926).

The minimal moral to be drawn from such observations is that the neural mechanisms that assign times of occurrence to external events are labile and can be affected by instructions and other factors. This should surprise only naive mechanists who imagine that the nervous system functions like a mechanical system of fixed characteristics which must always give the same output for the same stimulus input, mediating a simple immediate translation of stimulus impact into neural event and neural event into experiential element; but it should not surprise psychologists who accept that the nervous system has computational abilities and processes information.

Gestalt psychology conveyed a similar message. One dot on a page carries no implication of an extended pattern, but three placed at random define a triangle, or a straight line: A configuration emerges that was not previously present as its components, again confounding simple mechanists. *Gestalten* were not only spatial but also temporal. If one briefly exposed point is followed after a suitable short interval by a second point at a suitable distance, we see only a single point that moves smoothly between the two locations (Ternus 1926/1938). This is a *temporal* gestalt. What it shows is that the neural computation of movement can take account of patterning of input over a short period of time, and give a correspondingly complex output.

Evidently such a computation cannot conclude before the second point has been received. For how long a period can the conclusion of a neural computation be delayed? Light is thrown on this question by an experiment of Treisman and Howarth (1959). Earlier work had shown that if the threshold for detection of a faint stimulus (the critical, to-be-detected stimulus) is measured, this threshold is lowered if a preceding accessory or "warning" stimulus is given that conveys information about the time at which the critical stimulus will subsequently be presented. This threshold lowering may be seen even when the critical stimulus is delayed four seconds or more after the accessory stimulus (Treisman 1946). Treisman and Howarth (1959) extended this paradigm to examine the effect of an accessory stimulus that *came at a fixed interval after* the presentation of the critical stimulus. Figure 1 shows the results of one such experiment. Here the subject listened to a constant 500 Hz tone presented over earphones at 60–70 dB SL. The stimulus to be detected was a 50 msec increment in the intensity of the tone. A prewarning (the onset of a neon light which came on and stayed on) initiated the trial. After a randomly chosen interval of 2–7 seconds, a near-threshold intensity increment was presented. In the warning conditions, this was followed after a fixed interval known to the subject (3, 1, 0.5, or 0 seconds) by

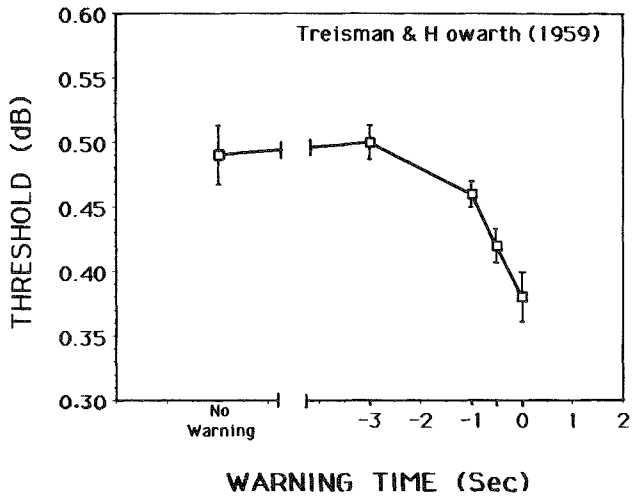

Figure 1 (Treisman). Data from Experiment 3 (b) of Treisman and Howarth (1959). The threshold in decibels for detection of an intensity increment added to a constant tone is plotted against Warning Time in seconds, where a negative Warning Time indicates that the accessory stimulus was presented later than the intensity increment. Standard error bars are shown.

presentation of the accessory stimulus, the onset of a second neon light which stayed on until the subject responded.

The data show that the threshold for detecting the intensity increment is lowered when the onset of the accessory stimulus occurs at the same time as the increment is given, or even when the accessory stimulus follows the critical stimulus by as much as one second. This effect was not accompanied by any increase in variance or in false positive rate (Treisman & Howarth 1959). This demonstrates that the neural sensory mechanisms can store sensory input information for a period of up to one second, and can use subsequent contextual information received during that delay to revise the criterion that is eventually applied to determine the detection decision regarding the stored sensory information. This stored information is not "conscious," to the extent that this is indexed by verbal report: If the subsequent warning is not given, the subject will deny hearing any increment on a proportion of trials on which, if the warning stimulus had been given following the critical stimulus at the expected interval, a positive response would have been made. Thus, not only are the sensory neural mechanisms capable of storing and "editing" information, but the length of this storage can be of the order of as much as a second.

It is hardly necessary to note that many types of observations preclude such delays being obligatory and universal. We can safely conclude that the extent of the delay before the neural mechanisms reach a perceptual decision on an input may vary with the context in which the subject operates and the nature of the task.

The experiments discussed by D & K report apparent motion between two differently colored shapes, in which the color that is perceived changes during the course of the motion (Kolers & von Grünau 1976); apparent movement of mechanical or electrical cutaneous pulses given successively at different positions on the skin (Geldard & Sherrick 1972) – both of which are temporal gestalts; and two experiments in which subjects reported "the subjective timing order of two sensory experiences" (Libet et al. 1979; Libet 1985a; 1985b).

At the neurophysiological level, there is nothing puzzling about these results. The Kolers and von Grünau experiment adds nothing of theoretical significance (for the present purpose: the mismatch between perceived timing and a naive mechanistic account of its basis) to earlier reports of apparent motion: When a point is seen moving from one position to another, the second position must be defined by sensory input before the neural mechanisms can determine the direction of movement. What any observation of apparent movement or displacement tells us is that the processing capacity of the neural mechanisms that determine the eventual perception is sufficient to allow the conclusion of the processing to be timed to accommodate the physical delays embodied in the complex stimulus. This point is also relevant to Libet's observations, although the standing of his data is somewhat vitiated by serious methodological problems. He notes (Libet 1985a): "We tended to place the subjective experience at the times reported by the subject, these reports being the only available directly valid indicators of such times" (p. 564). "Validity" can only be claimed for a measure of a variable when that measure can be compared with and shown to agree with an accepted objective indicator of the value of the variable we wish to measure. There is no way of doing that here: The validity claimed is based on hopeful thinking. When the subject must report what he believes the clock position was at the time he believes he observed in himself a conscious intention to act, we are back with the prior entry phenomenon, with its known susceptibility to internal and external factors. The belief that various electrophysiological measures provide information about the timing of conscious experiences also cannot be validated: It depends on the interpretation of the content of verbal reports.

Leaving these problems aside, what we have are situations such as the subject reporting a clock position and an intention to act and the experimenter recording a cortical potential; and the constellation of observations and reports in such a situation can be attributed to a corresponding constellation of neural processing. Without a semantic analysis of the subjects' reports, no metaphysical problems arise, as D & K appreciate (sect. 1.2, para. 9).

I suggest we can take this further. Even if the semantic content of subjects' reports is taken into account, none of the experimental observations above will present any paradox if we accept the following two points:

1. Neural information processing does not simply mirror the physical input to the senses. The neural state that is the response to a complex of stimuli is not simply the superposition of the neural states that would be produced by the component stimuli presented alone. The pattern of the stimuli (in space or time, subject in each case to limits) determines the specific complex neural response. This is not a novel point: it was made by the gestalt psychologists and is contained in almost every theory of neural information processing.

2. We take it that Relation_u guarantees a consistent and unvarying association between states of conscious awareness and neurophysiological states; for example, Relation_u might be taken to be an identity or similar relation.

If these points are accepted, no paradoxes arise from the contents of subjects' reports because the constant relation between the relevant complex neural activity and what D & K call "the subjective sequence of the stream of consciousness" (sect. 1.2, para. 10) means that there can be no information in the latter that is not contained in the former. All that can be transferred from the neural computation to its conscious correlate is information, and the constant relation hypothesis implies that there is no additional information in the latter, generated in some immaterial manner. Speech cannot convey more than information, thus the information in subjects' reports must correspond equally well to the information manifesting in or as "conscious experience" and to the same information computed neurally. Thus nothing in a subject's report can be paradoxical in the sense of implying something not possible physically, since it can imply no more than the information content of the neural computation giving rise to it, and (it is generally accepted that) neural activity complies with the laws of physics.

This is not to say that experimental reports such as those

discussed above may not cause surprise, but the surprise is that of the naive mechanist forced to acknowledge that in the nervous system he faces something more complex than a simple mechanism.

This explanation for these results does not accord with the lines taken by D & K. What leads them to their different view? Basically, it would seem, mistaking the details of poor models of neural processing for necessary conditions on such activity.

D & K accept that paradoxes seem to arise from observations of the perception of temporal sequence, and compare "two models of consciousness" that they consider have a claim to explain them. First,

> According to the standard "Cartesian Theater" model, there is a place in the brain where "it all comes together" and the discriminations in all modalities are somehow put into registration and "presented" for subjective judgment. The timing of the events in this Theater determines subjective order. (target article, Abstract)

As an alternative to the "standard Cartesian Theater model" (CT model) they put forward a "Multiple Drafts" (MD) model. According to this,

> All perceptual operations . . . are accomplished by multitrack processes . . . discriminations *have to be made only once* . . . what effect [a localized discrimination] has on the prevailing brain state (and thus awareness) can change from moment to moment. . . . Drafts of experience can be revised. . . . These content-fixations are themselves precisely locatable in both space and time, but their onsets do *not* mark the onset of awareness of their content. It is always an open question whether any particular content thus discriminated will eventually appear as an element in conscious experience. . . . at any point in time there are multiple "drafts" of narrative fragments at various stages of "editing" in various places in the brain. (sect. 1.1, para. 10)

Unfortunately, D & K do not use a precisely defined vocabulary to describe their model, and this creates ambiguities. It is not always clear whether terms such as "content" or "discrimination" refer to neural processes or to subjective phenomena or have an ambiguous double life in both spheres. Attempting to resolve these ambiguities, I take it that what the MD model claims is

1. Afferent information from the sense organs undergoes neural processing, which extracts information. This processing occurs at various parts of the brain. The ongoing neural processing is continuously modified by the impact of further incoming sensory information. (Neural processing in general, and the incorporation of additional information in such processing is what I take the metaphors of "drafts" and "editing" to refer to.)

These assumptions are compatible with most accepted accounts of the neural computation underlying continuing action in the real world, such as the performance and control of skilled movements. A memorable illustration of this point was given by Lashley (1951) in his example of the spoken sentence "Rapid righting with his uninjured hand saved from loss the contents of the capsized canoe": He noted that the spoken word "righting" (which may be heard at first as "writing") receives its final meaning 3 to 5 seconds after it is first heard.

2. "Contents," which I take to mean information extracted by neural information processing, may or may not "appear as an element in conscious experience."

3. The "stream of consciousness" . . . is a parallel stream of conflicting and continuously revised contents, no one narrative thread of which [is] the true version of conscious experience" (sect. 1.1, para. 11).

What basis do these views rest on? D & K reject "Descartes' interactionistic dualism, with its systematically inexplicable traffic" between two realms, and note that "materialism . . . is now a received opinion approaching unanimity" (sect. 1.1, para. 8). But it seems to be a materialism that allows "many commentators" to feel "That one's consciousness might lag behind the brain processes . . . seems . . . unsettling . . . ruling out a real "executive role" for "the conscious self" (sect. 1.2, para. 8); and that is compatible with "the *prima facie* plausible thesis that our conscious perceptions are caused by events in our nervous systems, and our conscious acts, in turn, cause events in our nervous systems that control our bodily acts" (sect. 1.2, para. 12).

But a more concrete vision moves them, a vision of the brain, extended in space, continuing through time, in which "representations" of various stimuli follow their erratic courses (like the Wandering Rocks in the *Odyssey*) perhaps eventually provoking action. How in this fuzzy context can order of arrival be determined? D & K see it as a "*logical* difficulty" that "if A beats B to one finish line but B beats A to another, which result fixes subjective sequence in consciousness" (sect. 1.1, para. 3)? D & K's answer is to deny the existence of a unique sequence: Different questions will evoke different answers. No one answer is canonical.

There is no "logical difficulty." The conclusion D & K derive from this vision is fallacious. It is based on what I consider to be two errors. The first is that they confound different meanings of "represent." In one sense this word refers merely to causal consequences: Weakness may represent a result of malnutrition. Firing in the afferent nerves from the knee, in the spinal cord, in the motor nerve to the quadriceps, and the final knee jerk, all represent consequences of a tap on the patellar tendon. But none of them is a "representation" of the tap. A stimulus does not produce a "representation" that travels through the nervous system. Nothing in that sense travels. Neurons fire in complex patterns, and this firing is modulated by the causal consequences of continuing sensory stimulation.

A "representation" in the second sense is a signal or event which has been assigned the burden of carrying a certain meaning: A skull and cross-bones may represent danger; a collection of letters may represent a spoken word; a painting may represent a scene. In this way, some pattern of firing at a limited or extended neural location could carry the meaning that a certain stimulus pattern exists in the external world. This neural activity could be temporally and spatially restricted or localised. It should not be confused with the complete set of neural consequences caused by the original peripheral stimulation.

The second error is the assumption that "subjective sequence" is determined "in consciousness." I believe it is determined by neural processing, just like any other aspect of the stimulus, such as color, shape, or location in space. The processing of the location of a sound-source in auditory space provides an interesting example. The location of the source determines the time of arrival of sound waves at each ear. Each auditory input has multiple neural consequences that almost certainly include specific neural messages directed to a locus in the midbrain at which the difference between their times of arrival is determined, giving a measure which further neural computation uses to determine the azimuth of the sound-source. We may be conscious of and able to report the computed direction in space of the sound-source but we are not conscious of the time disparity measure itself.

It is perfectly possible that such neural processing provides a valid model for analogous neural processing used to compare the times of arrival of different stimuli over longer intervals to provide a basis for overt judgments of sequential order. The point here is that it is not a matter of different "representations" being seen to arrive (by what implicit observer?) at an arbitrary "finish line" in consciousness. It is sufficient to establish that it is coherent to propose, as a neural model, that specific neural messages, engendered in relation to the action of given stimuli on the sensorium, are directed to a dedicated temporal locus of comparison which extracts relative times of arrival, just as other messages from the same stimuli may travel to areas which determine color or analyse binocular disparity. All that is required for such processing is an internal clock (Treisman 1963; 1984; Treisman et al. 1990) and it is perfectly possible that such a

mechanism could bias its measures to take account of whether the stimulus timed acted on the foot or the forehead.

On this understanding, D & K are led astray by failing to distinguish between neural processing on the one hand, and hypothetical processes proceeding "in consciousness" on the other, and their "model of consciousness" solves a nonexistent problem. Their belief that it is impossible to establish temporal sequence uniquely is a consequence of a mistaken model, not a necessary limitation in information processing.

A related difficulty is D & K's reluctance to accept the possibility of delayed processing. If the form of a temporal gestalt depends on the later stimulation "would the brain always delay response to one tap in case more came? If not, how does it 'know' when to delay" (sect. 1.2, para. 4)? D & K note "the biological implausibility of such a squandering of time" (sect. 2.2, para. 14). But it is not a question of plausibility; it is a question of what the facts are. The answers to these questions are given by the experimental work discussed above, which shows that in certain situations quite extensive delays may occur before a perceptual computation is concluded. These delays may reflect task requirements, and the subject's experience of the situation.

D & K attempt to solve these problems in a way that will avoid assuming such delays, and their explanation has implications they have not adequately considered. They reject the explanations for apparent motion which they attribute to the Cartesian Theater (CT) model, which apparently involve surgery on the contents of conscious experience, evoked by subsequent information. They offer instead a Multiple Drafts (MD) account in which, when two flashes are presented in succession each is discriminated, each presumably providing a corresponding phenomenal experience. Where then does the movement come from? When does it arise? According to the MD model: "Retrospectively the brain creates the content (the judgment) that there was intervening motion, and this content is then available to govern activity and leave its mark on memory. But . . . the brain does not bother 'constructing' any representations that go to the trouble of 'filling in' the blanks. . . . The judgment is *already in*, so the brain can get on with other tasks!" (sect. 2.2, para. 24).

This evades the difficulty of later phenomenal experience adjusting earlier phenomenal experience, whether by insertion or overwriting. But it fails to explain why, if we *phenomenally experience* two flashes, and *judge* that there was movement between them, we *see one* flash (not the two that on the MD account we phenomenally experienced). The MD claim is that we have a mosaic, two flashes that are experienced in a phenomenological sense, and movement between them that is not so experienced but is actually inferred and misremembered as having been experienced. This is a strange outcome for a theory that sets out to avoid "extravagances," not least because this account does not match what subjects report, whereas this debate rests heavily on the authority given to subjective report.

But this account must seem even stranger if we consider that apparent movement is not a rare phenomenon. When we watch a cinema film we phenomenally experience projected frames 48 times a second, and "just conclude that there was intervening motion" (sect. 2.2, para. 23) (of the many different directions and speeds required by the many constituents of each frame) and lodge this in memory, 24 times a second. Thus D & K have perception (of form) racing along in its naive mechanist way to keep in time with the stimulus input, frame by frame (avoiding biologically implausible delays), while consciousness limps along behind (conscious processes are allowed to tarry) "just concluding" what the motion was, at this fantastic rate per second. What a complex theoretical price to pay to preserve the naive mechanistic assumption that two successive flashes must give two phenomenal experiences, and to avoid accepting that it is within the powers of neural mechanisms to lag sufficiently behind input to give the best integrated interpretation of it.

D & K's view that "fixing the exact time of onset of some representing element in some place in the brain does not provide a temporal landmark" (sect. 2.1, para. 7), that is, there is no "fact of the matter about exactly when . . . a conscious experience happens" (sect. 3.1, para. 12) – raises the problem of how different trains of sensory stimulation – such as the auditory and visual consequences of external events – are matched up in perception. To solve this problem they propose a "biologically . . . plausible" model of "content-sensitive settling" (sect. 2.1, para. 8). This is based on an analogy with the process of synchronizing an audio tape with the corresponding length of film by finding convergences between events recorded in each. But this implies that the perception of these multimodal events must be determined *after* they have been recorded on the different media and the media then matched up, which seems incompatible with their stance above; and it overlooks that accurate time recording in the construction of each tape is necessary for such a process to work: Two such tapes could only be matched up if timing devices have determined the rate at which each was recorded. If the rate at which the film camera recorded frames and the speed of the tape-recorder motor were not fixed in a known relation, we would be unlikely to be able to find a match between patterns of salient events on the two records.

The criticisms above are those that appeal to a psychologist: failure to appreciate the potentialities of the neurophysiological mechanisms underlying processing; and the misconceptions that result if conclusions based on introspection are projected onto the neural processing mechanisms. But they do appear considerably to weaken the basis for Dennett & Kinsbourne's philosophical superstructure.

Time for more alternatives

Robert Van Gulick

Department of Philosophy, Syracuse University, Syracuse, NY 13244-1170
Electronic mail: *rnvangul@sunrise.bitnet*

Dennett & Kinsbourne (D & K) argue in favor of their Multiple Drafts (MD) model of consciousness and against what they call the Cartesian Theater model by showing that the former is able to explain various facts about the subjective timing of experience that remain anomalous on the latter. They make a convincing case against the Cartesian Theater, but they are less than clear about just what the MD model entails and about exactly which aspects of it are supposed to be supported by their arguments.

There are four main sorts of problems.

1. To what extent is the Cartesian Theater a straw man whose defeat confers little credit on the MD model? That is, what other alternatives are there?

2. In particular one can accommodate the temporal data on a model that is not so antirealist or indeterministic about the stream of consciousness?

3. Are there possible sources of evidence to decide questions about the content of conscious experience beyond those that D & K consider?

4. Can the phenomena to which D & K appeal support the general conclusions they want to draw or are they special cases of only limited relevance?

Let me deal briefly with each of these in turn.

D & K intend the Cartesian Theater as a model not only of historical Cartesianism but also of the Cartesian materialism they find implicit and widespread in current thinking about the mind. The theater metaphor is that of a place where "it all comes together," a materialist version of the Cartesian bottleneck between mind and body. The notion of *place* is understood quite literally as a spatial location; the view as D & K put it is that if we knew *when*, on the inward pathway of the stimulus, experience occurred, then we could say *where* it occurred, and vice versa.

Whatever the historical facts, it seems unfair to saddle current friends of consciousness with the view that it occurs at a distinct spatial location in the brain. The distinction between conscious and unconscious processing can be significant and fairly clear even if consciousness does not occur at a special place. Conscious states differ from unconscious ones in the processes they involve, not in where they occur. Consciousness may well be a more or less global brain state involving the simultaneous activation and interaction of many different brain regions and systems of representation; being spatially "smeared" in D & K's sense does not rule out a clear divide between conscious and unconscious stages of processing and awareness. Other features of the Cartesian Theater seem to unburden it unfairly also. As D & K present it, the model requires an observer who is all-seeing and all-knowing regarding the perfectly determinate projections on the screen of the inner theater. What is projected there can never be fuzzy or indeterminate, and though the observer may succumb to rapid loss of memory, he can never fail to notice anything on the screen before him at the instant of its presence. Such strong commitments make the Cartesian Theater an easy model to defeat, but again it is not clear that those who want to be more realist about consciousness than the MD model allows need to buy into such strong commitments.

Thus, turning to our second point, what might an alternative model look like? One possibility is to distinguish as conscious those representations and states of awareness that have phenomenal properties and structure; my present conscious visual awareness of the computer screen on which I am writing would be a good example. Such states need not be locally realized; indeed, the rich informational structure of phenomenal representation (e.g., when I am visually aware of the computer I am also aware of what it is and of its myriad relations to other sorts of items) would probably require the simultaneous interaction of many brain regions and representational systems. Since such states would be the global, integrative of activity in many subsystems, and would have phenomenal properties, we can call this alternative the Global Integrative Phenomenal State (or GIPS) model of consciousness. Like the MD model it denies that there is any central homunculus or any special location in which consciousness takes place. But unlike the MD model, it takes seriously the distinction between phenomenal and nonphenomenal representation; not all drafts and not all representations or content fixations count as conscious (nor is the difference just a matter of whether they can be reported, a criterion that would make all the mental states of nonhuman animals nonconscious). On the GIPS model, questions about a person's state of consciousness at a given moment are questions about the content of her phenomenal representations.

Can the GIPS model handle the data about subjective timing without anomaly? I think it can. Consider the color phi. The GIPS explanation would be more or less of the sort D & K call Stalinesque. Apparent motion in phenomenal consciousness would result when the interstimulus interval was shorter than that of the integrative interactions producing the phenomenal representation (note that representation need not be a completely independent downstream product of earlier representation as opposed to being a stabilized cooperative activation of them) The resultant lag (of 200 ms or less) need not be biologically implausible, as D & K claim given

1. that it is so important to be aware of motion,
2. that the relevant integrative processes (not unlike what D & K call content-sensitive settling) will take some time, and
3. that it is possible to make automatic or reflex responses prior to and in the absence of conscious phenomenal awareness, as D & K themselves note.

The third of these facts can also explain why the lag or delay is not manifest in reaction time tests. Since the role of phenomenal awareness does not seem to be to trigger automatic responses but to provide a representation capable of planning and guiding flexible and variable responses, a delay of a few hundred milliseconds to carry out the integration generating such representations need not have any costly consequences.

The GIPS model would also favor the standard or Stalinesque explanation of meta-contrast or backward masking experiments. The mask prevents the initial brief stimulus from being integrated into phenomenal representation. D & K argue that the Orwellian and Stalinesque accounts disagree about "a difference that makes no difference." Their position is based on a quasiverificationist claim that the two versions can equally account for all the imaginable data. But (coming now to our third point), it seems that there could be evidence relevant to settling the dispute. D & K consider first person introspective evidence and third person evidence of verbal and nonverbal behavior. But there is also the possibility of third person neurological evidence, which at least in the future might be evaluated in terms of a theory of the neural basis of phenomenal experience that could tell us whether or not the masked stimulus even briefly generated what we independently knew to be the neural correlate of phenomenal representation of the relevant stimulus. Utopian hopes? Perhaps so. But in the absence of any principled reasons to rule out such a possibility a priori, it seems premature to conclude that the dispute is devoid of content.

In their summary D & K acknowledge that some critics may accuse them of overgeneralizing from special cases (the fourth sort of problem). And they admit with their mental counting example that in typical macroscopic cases our experiences occur in the temporal order in which we experience them to occur. They nonetheless claim that the special cases are representative of the brain's normal manner of functioning. However, the reasons they cite in favor of this claim seem somewhat beside the point. They appeal to the fact that our memories of past events tend to be in the form of internally coherent simplified descriptions that frequently lose or misrepresent details of temporal sequence. This is undeniably true and it is consistent with their view of the mind as continuously trying to spin consistent and coherent scenarios, but it leaves untouched the central claim that in our *perceptual experience* of events (as opposed to our memories of such events) we experience the sequence of events by experiencing them in that sequence. The claim of Mellor's (1981) that D & K dispute is entirely about the *perceptual experience* of succession, and it seems untouched with respect to macroscopic perception. Facts about memory consolidation seem irrelevant to the issue.

Thus I don't think D & K can be said to have dislodged or dealt adequately with the view that phenomenal consciousness involves what Kant called intuitions, i.e., continuous sensuous manifolds of time and space within which phenomenal objects and events are presented. Kant and a multitude of others may have been wrong in supposing that conscious experience requires such manifolds, but that remains to be shown. The GIPS model draws a clear distinction between phenomenal and nonphenomenal representation, and thus in the end it must provide some account of the structure of phenomenal space and time. The Multiple Drafts model as I understand it would give little or no weight to the phenomenal/nonphenomenal distinction and might thus avoid giving any such account. In order to avoid that burden, however, it must make its case for rejecting the distinction and I don't think that case has yet been made.

Is consciousness integrated?

Max Velmans

Department of Psychology, Goldsmiths College, University of London, London SE14 6NW, England
Electronic mail: *mlv@gold.lon.ac.uk*

In the visual system, the represented features of individual objects (shape, colour, movement, and so on) are distributed both in space and time within the brain. Representations of

inner and outer event sequences arrive through different sense organs at different times and are likewise distributed. Objects are nevertheless perceived as integrated wholes – and event sequences are experienced to form a coherent "consciousness stream." In their thoughtful target article, Dennett & Kinsbourne (D & K) ask how this is achieved.

According to Descartes (1644/1972), integration requires a single interface between conscious experience and the brain, provided by the pineal gland. Its central location is well suited to transforming volitions originating in the soul into movements and it provides a suitable point of convergence where disparate retinal images can be combined into a single, integrated visual field. In current theorizing the status of the pineal gland is less exalted, but it is still widely assumed that information arriving at the various sense organs is somehow *integrated* (with stored knowledge, needs, intentions, etc.) not only to allow co-ordinated, adaptive response, but also to provide an integrated conscious stream (e.g., Baars 1989; Dixon 1981; Marcel 1983; Navon 1991; van Gulick 1991; Velmans 1991b). A few theorists follow Descartes in suggesting a central co-ordinating system (for example, Dimond 1980; Penfield 1975); others allow that the integrated neural correlates of consciousness may be widely distributed throughout the brain (e.g., John 1976; Koch & Crick 1991; Pribram 1971; 1986).

According to D & K, these models subscribe to a form of Cartesian Theatre in that they assume human information processing relating to any given even to present a final, integrated representation (of that event) to the "footlights" of consciousness. Other than their commitment to cerebral integration, however, there is very little that is Cartesian about these views. No current theory adopts Descartes' neurophysiology or his account of processing. So it might be more accurate to call this the "integrationist" position.

In D & K's Multiple Drafts model there is *no* integrated "definitive stream of consciousness" in which information about the world all comes together; there is only a "parallel stream of conflicting and continuously revised contents." They claim that this avoids the "scientific and metaphysical extravagances" of the Cartesian Theatre, which assumes (1), that there is a *place* in the brain where information from all relevant inputs is presented for a final *subjective judgement* and (2), that the temporal ordering of experienced events reflects the temporal ordering in which representations of those events arrive in the theatre. By contrast, representations in the Multiple Drafts model are (1) distributed both in space and time in the brain, and (2) "a product of the brain's interpretational processes, not a direct reflection of events making up those processes" (Abstract). This characterization of current integrationist views, however, is very misleading.

As noted above, only a few supporters of cerebral integration propose that this occurs in a specific *place* within the brain, and even fewer would assume that integration requires prior *subjective judgement* (one exception night be Eccles 1980). Indeed, with respect to *processing* there is very little in D & K's description of the Multiple Drafts model that differentiates it from an integrationist account. Few integrationists would deny that representations are distributed, both in space and time within the brain; nor would they take issue with the view that information encoded *within* neural representational states may enter into subsequent mental functioning, rather than the physical properties of such states (cf. Uttal 1978; Velmans 1991b, note 30). D & K accept the integrationist assumption that in order to function with distributed representations, "the brain must be able to 'bind' or 'correlate' and 'compare' various separately discriminated contents" (sect. 1.1, para. 6). To allow integration, interpretational processes must combine information *about* inputs (contained *within* distributed representations) with stored contextual information in "real time" (cf. Blumenthal 1977). Indeed, few integrationists would deny the existence of "multiple drafts." The stream of consciousness might be integrated, but it is in continuous flux and change. It is difficult to envisage how dynamic change could be represented *without* constant redrafting of representational states. Even in the Cartesian Theatre the show must go on.

Consequently, the only difference between the models that "makes a difference" is whether or not an integrated conscious stream, with its corresponding integrated neural correlates, *exists*. This is the core of what D & K deny.

Faced with the obvious – that we normally experience just *one* world, not multiple worlds, and that we normally do so *without* conflict about what we perceive – they plead massive cortical self-deception! In the fraction of a second between multiple, conflicting "drafts of experience" and *subjective reports*, the "Orwellian" brain continuously rewrites history (on the basis of the latest available information) to produce a consistent "party line." Hence, we *report* having integrated experiences, although there are no *prior* integrated experiences to report. According to D & K, this scenario is more plausible than the "extravagant" integrationist alternative – a "Stalinesque" doctoring of the conflicting evidence *prior to* its presentation before the "show trial" of conscious experience.

The integrationist model, however, requires neither doctoring of the evidence nor adherence to any given party line. On the contrary, it usually assumes the brain to be making the most accurate mental model it can of what is going on, based on the best available information, stored knowledge, and so on, at the time of integration. As D & K ably demonstrate, relevant information may arrive at different times, even that relating to temporal sequencing. Provided that the information does not arrive too late, information *about* temporal sequence (rather than the arrival time of the information itself) is integrated into the way temporal sequence is experienced (thus accounting for Libet's "backward referral" findings). Colour phi and the cutaneous rabbit provide persuasive *demonstrations* of the way cerebral processes integrate what is often partial information, into a coherent, integrated experience.

Contrary to D & K's claim, this integrationist model is theoretically *less* "extravagant" than their Multiple Drafts account. It assumes that within any given time window, information is integrated into a single conscious stream and that unless there is evidence to the contrary, subjects' reports about what they experience are accurate. D & K require us to believe that subjective reports are *not* reports of prior experience; for example, subjects' claims to have experienced colour phi *prior* to reporting it are based, they suggest, on rapid forgetting of parallel, conflicting contents. It is what subjects *believe at the time of report* that defines "what the subject was conscious of" (sect. 2.2, para. 36).

This is theoretically self-defeating, for the reason that, at the time of report, subjects believe that they are describing a *prior*, integrated experience. A theoretical account that denies the accuracy of this belief is in danger of being unfalsifiable. Unless the conditions under which rapid forgetting occurs are precisely stated (cf. Velmans 1991b, Note 6), one can invoke it at will to deny the legitimacy of *any* theoretically inconvenient reports. In short, Dennett & Kinsbourne can be accused of doctoring the evidence. It is their argument, rather than the integrationist model, which is Stalinesque.

Global pattern perception and temporal order judgments

Richard M. Warren

Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI 53211

Electronic mail: *rmwarren@csd4.csd.uwm.edu*

I agree with Dennett & Kinsbourne (D & K) that events occurring in the brain "do have temporal properties, but those

properties do not determine subjective order." I also agree (with qualifications given below) that "we perceive – and remember – perceptual events, not a successively analyzed trickle of perceptual elements or attributes locked into succession as if pinned into place on a continuous film" (Conclusion, para. 3). However, since I have reached these general conclusions through a rather different route, my formulation of basic principles is somewhat different.

The conclusions drawn in the target article are based largely upon analyses of illusions and apparent paradoxes involving successive events, which are called "some 'temporal anomalies' of consciousness" (sect. 1.2). I believe that these anomalies can be explained most readily with the help of two principles enunciated by Helmholtz during the last century: (1) We are unable to observe our sensations accurately except as they are useful in enabling us to recognize external objects; and (2) illusions are the consequence of relating our sensory input to objects (or events) which would normally produce a similar input (Warren & Warren 1968, pp. 129, 175–76). Both the "cutaneous rabbit" (sect. 1.2B) and "color phi" (sect. 1.2A) can be considered as an interpretation of sensory input in terms of events which would produce similar patterns.

Recall that the cutaneous rabbit appears to be a single saltatory object (the hopping rabbit) traveling up the arm in equidistant steps, despite the fact that the stimulus consists of sequences of taps delivered in turn to each of three fixed positions on the arm (wrist, near the elbow, upper arm). On catch trials when only the series of taps to the wrist are given, the rabbit hops in place. Yet, if this same succession of stationary wrist taps is followed by elbow taps, the rabbit is perceived to move after the very first wrist tap. This observation loses its paradoxical flavor if we consider that a particular series of repetitive taps is interpreted not as a succession of independent events (taps), but rather as an integrated pattern produced by a single causative agent. This agent cannot readily jump abruptly from one of the three stimulated positions to the next within a single inter-tap interval (which ranged from 50 to 200 ms).

The classical phi illusion is produced by two stationary lights separated by a few degrees: One of the lights is on while the other is off, each flashing at a rate not too far below the flicker fusion frequency. This display resembles that produced by a single light source which is lit continuously and moves back and forth between two end points. Such a moving object would have the longest dwell time at the end positions as it slows to zero velocity and reverses direction. Midway between these positions, the velocity would be greatest, and the retinal image dimmest. The illusion of phi occurs when the neural response pattern approximates that which would be produced by such a moving object. In the modification of the classical monochromatic phi described in the target article (see sect. 1.2A), the flashing lights producing phi have different colors. The illusion of movement still takes place (and the explanation offered for the classical illusion still applies), with the change from one color to the other referred to a location equidistant from the two endpoints.

This alternative analysis of the illusions differs from that given by D & K, but is nevertheless consistent with their statement that "the temporal order of subjective events is a product of the brain's interpretational processes, not a direct reflection of events making up those processes" (last sentence of Abstract). However, I believe that it is not only "the temporal order of subjective events" that reflects interpretational processing by the brain, but that perception in general follows this interpretational principle. As an example of the broader application of this rule, I have suggested that when someone is asked to provide quantitative judgments of sensory intensity, the "brain's interpretive processes" evaluate input in terms of quantitative changes of some familiar physical scale associated in a regular fashion with changes in the level of stimulation. This approach to psychophysical scales has served as the basis for a target article on the measurement of sensory intensity in this Journal (Warren 1981).

To return to the topic of the temporal ordering of sensory events, experiments with auditory sequences have indicated that much of what appears to be direct perception of temporal order involves recognition of a familiar pattern, followed by a learned or inferred analytical description of components in their proper order. Since this hypothesis has implications not only for the "temporal anomalies" (sect. 1.2) described by D & K, but also for the way sequences of brief items are normally processed and interpreted, I will summarize briefly some empirical findings and conclusions based upon this research.

The threshold for direct naming of the order of items in iterated sequences of four sounds ranges from about 100 ms for sequences of vowels (Dorman et al. 1975; Thomas et al. 1971) to about 300 ms for a succession of unrelated items such as hisses, tones, and buzzes (Warren & Obusek 1972). The limiting duration for direct identification of order in such sequences appears to be the time required for attaching verbal labels to successive items as they occur, with the minimum time required for sequences of vowels – for which the name identifying the sound is the same as the sound itself (Teranishi 1977; Warren 1974a). This would appear to present difficulties for understanding speech, since the phonetic components have average durations below 100 ms. However, when listeners are charged with the task of *discriminating* between permuted orders of items, they can do so readily at all item durations down to at least 10 ms, not only for sequences of vowels (Warren et al. 1990), but also for sequences of tones (Warren et al. 1991) and unrelated sounds (Warren & Ackroff 1976).

These and related experiments led to the hypothesis that series of brief events are perceived holistically as "temporal compounds" having different ensemble characteristics with different arrangements. Although such compounds cannot be decomposed into their elements directly, it is quite easy for listeners to learn the appropriate verbal labels describing the components in their proper order (Warren 1974b). When this occurs, listeners are not always aware that their temporal analysis is inferential rather than direct.

Thus it appears that the identification of order within sequences of brief items involves a two-step process: (1) recognition of the patterns; and (2) recitation of a learned analytical description. It should be emphasized that while the identification of temporal order with brief events depends upon prior pattern recognition, recognition of temporal patterns does not require the identification of order, or even the identification of the items themselves (for a discussion of the application of this principle to speech perception, see Warren et al. 1990). However, D & K reverse the order of the two steps outlined above in their description of the basis for phi. They consider that, when two spots are projected successively at different positions on a screen, "If the brain determines 'first A, then B' the spot is seen as moving in one direction; if the brain determines 'first B, then A' the spot is seen as moving in the opposite direction." They go on to say that "this discrimination is, then, as a matter of logic, based on the brain's capacity to make a temporal order judgment of a particular level of resolution" (sect. 2.1, para. 12). I believe that if we consider instead that patterns are recognized first and then properties of these are patterns inferred, the conceptual difficulties involving temporal anomalies and paradoxes described by D & K are avoided.

In conclusion, although disagreeing on details, I agree with the basic position of Dennett & Kinsbourne that we perceive events rather than an ordered series of discrete elements. I hope that my commentary will be considered as furnishing support and extending the scope of this basic principle.

The psychoanatomy of consciousness: Neural integration occurs in single cells

Gerald S. Wasserman

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364

Electronic mail: *codelab@psych.purdue.edu*

There is much that is very attractive about Dennett & Kinsbourne's (D & K's) psychoanatomy. Not the least is the frank and forthright rejection of the shopworn notion that "the brain has any deeper headquarters, any inner sanctum" that mediates conscious experience. Flourens and Lashley would no doubt approve. Moreover, it was particularly interesting to read their scholarly discussion of the difference between what I had called (in Wasserman 1985, p. 556) "a measuring operation and the thing being measured." In D & K's terminology, this is the difference between "the temporal properties of representations" and "the representation of temporal properties." Finally, D & K offer an elegant and thorough presentation of the obligatory implications of the spatial and temporal dispersion of sensory signals in the brain. The intertwining of such dispersive effects with the possibility of conversion from temporal to spatial representation and vice versa (Uttal 1973) is fundamental to any understanding of mental function.

To this generally sound discussion, I would only add the caution that very little of the temporal "smear" in brain, ranging, according to D & K, "up to several hundred milliseconds," is due to axonal conduction delays. Even considering structures that are putatively distant from the brain (like hands), the conduction times of most axons are usually at least an order of magnitude smaller than the times that would be necessary to account for the temporal dispersion actually found in the brain. Much of the dispersion is due to the temporal summation of slow potentials inside single cells, of which more below. Hence, we should distance ourselves from the tendentious notion – promulgated in the work of Libet cited by D & K – that the brain is particularly far removed in time from the hand. It is not.

Certain particular features of the target article, however, are problematic. Most troubling is the fact that the justified ringing down of the curtain of the Cartesian Theater goes too far – to the point where the authors neglect the constraints imposed by the way neural integration actually occurs. D & K do recognize that integrative activities are "precisely locatable in both space and time" and that the "objective temporal properties of discriminating states may be determined . . . ," but they recognize no important consequences of these facts.

Yet there are consequences for, as far as is now known, neural integration only occurs in the brain in one way: by the summation of graded potentials inside single cells. In this regard, it is worth stressing that the behavioral data under discussion by D & K were elicited by experiments in which the relations of at least two stimuli were under investigation. For such stimuli to interact, the neural representations of the two stimuli must integrate. Despite certain interesting speculations (e.g., Penrose 1990, pp. 643–706), there is no action at a distance in brains. This means, as far as I know, that there is no way in which the representation of one thing can integrate with the representation of another thing *unless* those representations integrate in the literal sense of the word – by coming together. The only way this can be done is by both projecting to at least one cellular locus.

This does not mean that a single cell is any sort of theater observed by a homunculus; what expresses the outcome of integration in a projection neuron is the depolarization of its axonal pacemaker at the axon hillock; in an intrinsic neuron, what also matters are the depolarizations of the several presynaptic specializations of its dendritic arborization.

To be certain of not being misunderstood: The distributed anatomy of the brain and the consequent spatiotemporal dispersion of biological signals influences neural integration by determining what connects to what and when. Regional influences are also known: Activity-dependent accumulation of potassium in extracellular space modifies potentials across neighboring cell membranes, as do field potentials produced by current flows generated in neighboring cells. But the integration itself only occurs in a particular way and in very small places – by summation of slow potentials inside the membrane of single neurons. All these other influences only determine whether integration will occur and to what degree; they do not in themselves produce any integration.

It is because of this fundamental property of the brain that we have been studying the temporal dispersion of the cellular representations of sensory signals. We find that such cellular signals in natural neurons have properties that correlate with temporal aspects of perception (see the reviews given in Wasserman & Kong 1979; Felsten & Wasserman 1981; Nisly & Wasserman 1989; Wasserman et al. 1990). And we find that modifying the temporal waveform of such signals in artificial neurons influences human perception (Wasserman et al. 1990). The fact that such correlates exist in a single cell, of course, by no means implies the vulgar concept that conscious perception occurs there.

A second, less troubling, aspect of the target article is the overinterpretation of certain experimental data, particularly those on the color phi phenomenon reported by Kolers and von Grünau (1971). It is difficult for me to see why it matters at all, in the present context, that color phi is putatively discontinuous while shape phi is putatively smooth. However, if this claimed difference is of any theoretical importance, it should be recognized that the experiment had certain procedural characteristics that should lead to caution in accepting this characterization.

It will be recalled that Kolers and von Grünau did not rely on introspection, which we now know is vulnerable to Orwellian or Stalinesque distortion. Instead, they did an actual experiment – they measured the colors and shapes of perceptions exhibiting apparent motion at varying positions along their movement trajectory. They made these measurements by providing observers with stationary comparison stimuli which they could adjust to match their perception of the movement.

The measurements of shape phi and color phi, however, were not equally sensitive: In the former case, the apparatus was arranged so that the shape of the comparison stimulus could be adjusted to provide a completely satisfactory match to the shape of the moving perception for two of their three subjects and a fair match for the third.

But the control of color was considerably less satisfactory: The color of the comparison stimulus could *never* match anything any subject perceived. This limitation occurred because the test stimuli that evoked the apparent movement perception were produced by broad band light reflected from colored construction papers while the comparison stimulus was produced by narrow band light selected by a grating monochromator. Hence the saturations of the perceptions evoked by these stimuli were perforce always markedly different.

All the subjects could do to "match" the color of the moving perception was to vary the wavelength of the monochromator. Hence, they were judging the similarity of one quality, hue, in the presence of large differences in sensory quality. And they had no way of producing measurements that would indicate a graduated change of color appearance other than a change in dominant wavelength.

In hindsight, one can see that a satisfactory answer to the question now under discussion by D & K would have emerged had Kolers and van Grünau provided their subjects with the controls of a tristimulus colorimeter instead of just the single knob of a monochromator. (In fairness to Kolers and von Grünau, it should be noted [as D & K do note] that the principal question they had designed their experiment to answer was different from the one now under discussion: Kolers and von

Grünau wondered whether the hue of the moving perception would go across the color circle from red to green or around the color circle from red through yellow to green. Kolers and von Grünau's experiment quite satisfactorily rejected the latter alternative in favor of the former.)

I do not see that the smoothness of color phi matters very much, but even if there is no god in the brain machine, God is always in the details.

Closing the Cartesian Theatre

Andy Young

Psychology Department, Durham University, Science Laboratories, Durham DH1 3LE, England

Electronic mail: *a.w.young@durham.ac.uk*

When we look around us, we experience an orderly arrangement of objects in a three-dimensional environment. The apparent orderliness and unity of everyday experience can seem hard to reconcile with the evidence that perception is achieved through the deployment of functionally separable mechanisms for processing different types of information (and not just in terms of information represented in different sensory modalities; even within the visual system there is evidence of differences between the processing of shape, colour, movement, etc.). It is tempting to resolve this paradox by thinking that somehow, somewhere in the brain, there must be some form of higher-level integration.

Perhaps there is, but of course there is no need to integrate everything. Integration is needed for certain purposes and not others. As Dennett & Kinsbourne (D & K) make clear, we do not need to assume that consciousness reflects what is now showing at a central Cartesian Theatre. Yet, as the target article demonstrates, this notion has proved difficult to shake off, and still pervades much of our thinking. Until quite recently I had a season ticket too.

There are other reasons than problems in subjective timing for closing the Cartesian Theatre. The problems have become clear to me in considering the mounting evidence that brain injury can create different forms of selective loss of awareness (Schacter et al. 1988; Weiskrantz 1986; Young & de Haan 1990). For example, cases have been described in which brain-injured people can accurately localise visual stimuli they claim to be unable to see, can show evidence of recognising faces without awareness of recognition, can learn characteristics of visual stimuli they do not remember having seen, and so on. Nearly all such reports emphasise that there is no evidence of any overall, global change in consciousness, but rather a selective loss of certain aspects of normal experience. Although it is possible to accommodate these findings by thinking in terms of disconnection from a *single* conscious mechanism (Schacter et al. 1988), the sheer diversity of the phenomena has led some of us to wonder if this will prove adequate (Young & de Haan 1990). As Stone and Davies (in press) point out, the multiple dissociations revealed "make us realise that the phenomenon of mind is much more complex and heterogeneous than we may initially be disposed to believe." The Multiple Drafts model therefore looks more plausible.

Returning to the point that some things do need to be integrated, it is useful to consider an example; I will use lipreading. Understanding speech only by lipreading is very difficult, but there is now considerable evidence that lipreading none the less provides support for speech comprehension, even in people with normal hearing (Campbell 1989). To do this, the brain must integrate information specified in auditory and visual sensory modalities. An illusion discovered by McGurk and MacDonald (1976) shows this neatly. In this illusion, a mismatch between auditorily and visually specified phonemes results in the perceiver blending the two; when watching a video of a person mouthing the phoneme "ga," which is synchronised with the soundtrack "ba," most people *hear* the fusion as "da." Yet neither the visual nor the auditory recording track carries the signal "da"; it is a genuine fusion. The reason we possess this skill is not yet known with certainty, but it is likely that it relates to the demands of learning to decode speech in infancy, when a lot of time is spent watching people talking. Lipreading would then be particularly useful because some of the sounds which are difficult to distinguish auditorily are amongst those which are relatively easy to lipread. [See also *BBS* multiple book review of Massaro's *Speech Perception by Ear and Eye*, *BBS* 12(4) 1989.]

Like many illusions, the McGurk and MacDonald (1976) effect is very compelling. You can easily convince yourself what is on the auditory recording track by shutting your eyes, but as soon as you open them again, the audio-visual fusion takes over. Roberts and Summerfield (1981), however, have shown that if one measures selective adaptation to these fused stimuli (with eyes open), this is influenced *only* by what is on the auditory recording track, *not* by the fused percept we hear. At one level, this is entirely consistent with the Multiple Drafts model. One (auditory + visual) draft corresponds to what we report that we hear, but another (purely auditory) draft is the locus of the selective adaptation effect. However, the perceiver is not conscious of the existence of the purely auditory draft. Perhaps we are only meant to give something draft status if it is a conscious draft, but even if this is so, there are nonconscious contributions to be accounted for under whatever alternative name they are allocated.

This is an obvious point, but I think it is important. In accepting that the Cartesian Theatre oversimplified the problem of consciousness by giving it a single locus, we should not commit the opposite error of assuming that consciousness is everywhere in the brain. The idea of Multiple Drafts is a help, but some aspects of its relation to consciousness still need to be worked out.

Editorial Commentary

If there are Orwellian (misremembered) and Stalinesque (misperceived) events then it would seem that there must be unproblematic Reaganesque (veridical) ones too (such as the instant when I actually had my simple, punctate, "pink-elephant-now" thought) – otherwise, relative to what are the other two mistaken? Moreover, if the fact of a matter is neither objectively nor subjectively ascertainable, it is not clear why it should follow that it does not exist (e.g., the fact about whatever the true correlation – surely nonzero – might be between how things appear and how they are). Not only does there seem to be no reason to believe that the nonlocalizability of conscious events in real time is a metaphysical rather than just a methodological problem, but, as S. Sternberg (private communication) has pointed out, it may not even be a methodological problem in principle, only in practice: In principle, if I could have enough punctate pink-elephant-now thoughts in enough otherwise variable contexts, their neural invariants could be identified by sufficiently powerful techniques of signal/noise averaging and analysis. It may be like looking for a needle in a haystack, but surely there *is* a point in real time when at least my Reaganesque pink-elephant-now thoughts really occurred, even if their neural substrate was complicated and distributed. Problems with mental timing are manifestations only of the incommensurability (apart from quantitative correlation) of objective and subjective qualities, not the nonexistence or indeterminacy of the latter. As long as the Cartesian Theatre is not occupied by a homunculus, it seems a perfectly reasonable

setting for a Reaganesque melodrama experienced by *me*. And it's the absolute timing of my *experiences*, not my "representations" that's at issue, is it not? [Note that none of this has anything to do with (1) veridicality relative to any external object or event that the pink-elephant-now thought is "about" because the experience is completely endogenous. Nor is (2) the accuracy of memory relevant, because all we want to know is when the actually experienced *now* actually occurred. Likewise irrelevant is (3) the possibly distributed ("imperial") nature of the concurrent neural event that corresponded to *now*. And surely irrelevant too is (4) the signal-detection-theoretic platitude that at a sufficiently minute scale the punctate *now* experience, like all psychophysical judgments, turns out to be smeared across an interval of uncertainty rather than being mathematically punctate, because the "grain" of that smear is subexperiential.]

Authors' Response

Escape from the Cartesian Theater

Daniel C. Dennett[a] and Marcel Kinsbourne[b]

[a]*Center for Cognitive Studies, Tufts University, Medford, MA 02155;* [b]*Behavioral Neurology Unit, Sargent College, Boston University, Boston, MA 02215*

Electronic mail: *ddennett@pearl.tufts.edu*

We wish to thank the commentators for their largely constructive criticism. It is gratifying to discover that some of them had been thinking – and in some cases publishing – ideas along similar lines. We claimed that Cartesian materialism, the view of the brain with some deep center where "it all comes together" for consciousness, often seduces even those who explicitly reject it. As Neisser (1976), an early critic of the view, has said: "It is currently a very popular notion, and with good reason. It represents a theoretical coup: Not only are the facts of attention apparently explained, but psychology's most elusive target is finally nailed down to a box in a flow chart" (p. 103). Even now, **Damasio** remarks, it "informs virtually all research on mind and brain, explicitly or implicitly." Indeed, serial information processing models generally run this risk (Kinsbourne 1985). The commentaries provide a wealth of confirming instances of the seductive power of this idea. Our sternest critics (**Block, Farah, Libet,** and **Treisman**) adopt fairly standard Cartesian positions; more interesting are those commentators who take themselves to be mainly in agreement with us, but who express reservations or offer support with arguments that betray a continuing allegiance to one or another tenet of the view we sought to discredit.

The issues are extraordinarily slippery, and reading the commentaries has been a daunting experience. If only we had thought of putting it *this* way, rather than *that* way – it would have forestalled one all-too-reasonable objection or another. We first present, then, a refined summary of our central view, much improved, we think, by its annealing in reaction to the commentaries. Then we turn to the major themes expressed in the commentaries, and finally we respond to additional important issues raised.

Table 1. *Outline of Response*

1. Multiple drafts: an improved summary (**Roskies & Wood, Lycan, Glymour et al., Jeannerod, Libet, Lloyd, Baars & Fehling, Aronson et al., Block, Clark, Teghtsoonian, Treisman, Velmans, Shepard**)
2. Is there a fact of the matter? (**Lycan, Van Gulick, Glymour et al., Antony, Aronson et al., Lloyd, Velmans, Treisman, Clark,** EDITORIAL COMMENTARY, **Rollins**)
3. When does "filling in" happen? (**Shepard, Roskies & Wood**)
4. Consciousness as a "system" with a function (**Antony, Teghtsoonian, Baars & Fehling, Velmans, Farah, Lycan, Reingold, Bridgeman, Lloyd, Block, Van Gulick**)
5. Libet and Treisman tilt at philosophers (**Libet, Treisman, Aronson et al., Van Gulick, Hurley**)
6. Other objections (**Wasserman, Gregson, Warren, Block, Van Gulick, Glymour et al., Shepard, Aronson et al., Lloyd**)
7. Supporting arguments (**Damasio, Hurley, McDermott, Reingold, Rosenthal, Young**)

1. Multiple Drafts: An improved summary

All the work that was dimly imagined to be done in the Cartesian Theater has to be done somewhere, and no doubt it is distributed around in the brain. This work is largely a matter of responding to the "given" by *taking* it – by responding to it with one interpretive judgment or another. This corner must be turned somehow by any model of observation. In the traditional view, all the taking is deferred until the raw given, the raw materials of stimulation, have been processed in various ways. Once each bit is "finished" it can enter consciousness and be appreciated for the first time. As C. S. Sherrington (1934) put it: "The mental action lies buried in the brain, and in that part most deeply recessed from the outside world that is furthest from input and output."

In our model, this single unified taking is broken up in cerebral space and real time. We suggest that the judgmental tasks are fragmented into many distributed moments of microtaking (Damasio 1989a; Kinsbourne 1988). The novelty in what we attempt here lies in how we develop the implications of this fragmentation. We have not merely broken the Cartesian Theater into thousands of minicinemas (**Roskies & Wood, Lycan, Glymour et al.**). Here is where the intuitive distinction between conscious taking and unconscious taking tends to beguile the theorist. It seems as if we are stuck with only three alternatives:

A. Each of these distributed microtakings is an episode of unconscious judgment, and the consciousness of the taken element must be deferred to some later process (the Stalinesque show trial in a Cartesian Theater). But then how long must each scene wait, pending potential revision, before the curtain rises on it?

B. Each of these distributed microtakings is an episode of conscious judgment (multiple minicinemas). But then why don't we all have either a kaleidoscopic and jumbled "stream of consciousness" or (if these distributed microtakings are not "co-conscious") a case of "multiple selves"?

Is our *retrospective* sense of unified, coherent consciousness just the artifact of an Orwellian historian's tampering with memory? As several commentators (**Baars & Fehling, Libet, Aronson et al.**) ask, how can the manifest coherence, seriality, or unity of conscious experience be explained?

C. Some of the distributed microtakings are conscious and the rest are not (e.g., **Block**'s "Cartesian modularism" with a spatially distributed "module"). The problem then becomes: What special property distinguishes those that are conscious, and how do we clock the onset of their activity (**Clark**)? And, of course, since distributed microtakings may occur slightly "out of order," what "mechanism" serves to unify the scattered microtakings that *are* conscious (**Teghtsoonian**), and in each case, does it operate before or after the onset of consciousness (i.e., which phenomena are Orwellian and which are Stalinesque)?

Our view is that there is yet a fourth alternative:

D. The creation of conscious experience is not a batch process but a continuous one. The microtakings have to interact. A microtaking, as a sort of judgment or decision, cannot just be inscribed in the brain in isolation; it has to have its consequences – for guiding action and modulating further microjudgments made "in its light," creating larger fragments of what we called narrative. In the target article we were silent on the mechanisms of interaction, but we did not mean to rule out processes that amount to partial construction of intermediate cases as an effect of a microtaking, rather than as its basis, as a way of ensuring the appropriate further influences (see our discussion of **Shepard** and **Roskies & Wood**, below). This interaction of microtakings, however it is accomplished in particular cases, has the effect that a modicum of coherence is maintained, with discrepant elements dropping out of contention, and without the assistance of a Master Judge. Because there is no Master Judge, there is no further process of being-appreciated-in-consciousness, so the question of exactly when a particular element was consciously (as opposed to unconsciously) taken admits no nonarbitrary answer.

Jeannerod provides an elegant demonstration of one prediction of this model when he and his colleagues demonstrate disunity of the self when examined in terms of timings of a fraction of a second. The time when the subject initiates a reaction to a stimulus depends on which effector – in this case hand or voice – is used, a finding incompatible with a single locus of decision in the Cartesian Theater (see also Marcel, in press). Whether they have really shown a dissociation between a preconscious (hand) response and a response based on a subjective experience (voice) is open to question. The amount of reprogramming needed to redirect a motor response already in formation would appear to be minimal and the reprogramming is accomplished under circumstances of very high stimulus-response compatibility. The vocal response, its relation to the stimulus being arbitrary, is less compatible. Possibly, had the task called for an imitative vocal response, subject to occasional minor modification, and an arbitrary manual response to such a change, the opposite difference would have been found: far briefer vocal than manual latency. There is no way of knowing whether the vocal response preceded, coincided with, or followed awareness of the stimulus. Indeed, there is nothing to know; all that can be said is that the two response types are directed by separate "drafts."

Some of the commentators (**Libet, Treisman, Velmans**) took us to be defending *B*, the Orwellian view. Patently we were not; we were instead demonstrating that it was the mirror image of *A*, the Stalinesque view, and that *neither* of them could evade the charge of unfalsifiability. **Glymour et al.** mistakenly supposed that our model "includes the notion of consciousness as an observer and interpreter of some 'draft.'" This is precisely not our view; interpretation is not reserved for some *one* draft; all drafts are products of (not just candidates for) interpretation, independently of consciousness. At least one commentator (**Lloyd**) took us to be defending *C* and criticized us for failing to provide the distinguishing mark of the conscious takings. Other commentators did understand that we were defending something like *D*, but they took it to be too radical, metaphysically. Because this is an almost overpoweringly plausible view, we will respond to it first, before going on to other objections.

2. Is there a fact of the matter?

The commentators generally agree with us that (1) the time of representing should not be confused with the time represented, and (2) there is no privileged place within the brain "where it all comes together." They do not all agree with us, however, that it follows from (1) and (2) that the Orwellian/Stalinesque distinction must break down at some scale of temporal resolution, leaving no fact of the matter about whether one is remembering mis-experiences or mis-remembering experiences. Here, some claim, we have gone overboard, lapsing into "verificationism" (**Lycan, Van Gulick, Glymour et al.**) or "eliminativism" (**Antony, Glymour et al., Aronson et al.**) or "antirealism" (**Lloyd, Van Gulick**), or some other gratuitously radical position (**Velmans, Treisman**). This is curious, for we consider our position to be unproblematically "realist" and materialist: Conscious experiences are real events occurring in the real time and space of the brain, and hence they are clockable and locatable within the appropriate limits of precision for real phenomena of their type. (For an extended defense of this version of realism, see Dennett 1991a.) Certain sorts of questions one might think it appropriate to ask about them, however, have no answers because these questions presuppose inappropriate – unmotivatable – temporal and spatial boundaries that are more fine-grained than the phenomena admit.

In the same spirit we are also realists about the British Empire – it really and truly existed in the physical space and time of this planet – but, again, we think that certain sorts of questions about the British Empire have no answers, simply because the British Empire was nothing over and above the various institutions, bureaucracies, and individuals that composed it. The question "Exactly when did the British Empire become informed of the truce in the War of 1812?" cannot be answered. The most that can be said is "Sometime between December 24, 1814, and mid-January, 1815." The signing of the truce was one official, intentional act of the Empire, but the later participation by the British forces in the Battle of

New Orleans was another, and it was an act performed under the assumption that no truce had been signed. Even if we can give precise times for the various moments at which various officials of the Empire became informed, no one of these moments can be singled out – except arbitrarily – as the time the Empire itself was informed. Similarly, since You are nothing over and above the various subagencies and processes in your nervous system that compose you, the following sort of question is always a trap: "Exactly when did I (as opposed to various parts of my brain) become informed (aware, conscious) of some event?" Conscious experience, in our view, is a succession of states *constituted by* various processes occurring in the brain and not something over and above these processes that is *caused by* them.

The idea is still very compelling, however, that "realism" about consciousness guarantees that certain questions have answers (even if they are currently unknowable). Embedded in the target article were grounds for rejecting this position but they did not stand forth for some of the commentators. Here, thanks to the light thrown by the commentaries, is a more succinct version of the basic argument.

The *only* difference between the Orwellian and Stalinesque treatment of any phenomenon is whether the editing or adjustment or tampering occurs before or after a presumed *moment of onset of consciousness* for the contents in question. The distinction can survive only if the debut into consciousness for some content is at least as accurately timable as the events of micro-taking (the binding, revising, interpreting, etc.) whose order relative to the onset of consciousness defines the two positions. If the onset of consciousness is not so sharply marked, the difference between pre-presentation Stalinesque revision and post-presentation Orwellian revision may disappear, and be restorable only by arbitrary fiat.

As "realists" about consciousness, we believe that there has to be something – some property K – that distinguishes conscious events from nonconscious events. Consider the following candidate for property K: A contentful event becomes conscious if and when it *becomes part of a temporarily dominant activity in cerebral cortex* (Kinsbourne, 1988, and in preparation). This is deliberately general and undetailed and it lacks any suggestion of a threshold. How long must participation in this dominance last, and how intense or exclusive does this dominance need to be, for an element to be conscious? There is no suggestion of a principled answer. Such a definition of property K meets the minimal demands of "realism," but threatens the presumed distinction between Orwellian and Stalinesque revisions. Suppose some contentful element briefly flourishes in such a dominant pattern but fades before leaving a salient, reportable trace on memory (a plausible example would be the representation of the first stimulus in a case of metacontrast). Would this support an Orwellian or a Stalinesque model? If the element participated for "long enough" it would be "in consciousness" even if it was never properly remembered (Orwell), but if it faded "too early" it would never quite make it into the privileged category, even if it left some traces in memory (Stalin). But how long is long enough? There is no way of saying (contra **Libet**, see below). No discontinuity divides the cases in two.

But perhaps we are overlooking some source of discontinuity; perhaps there is a sharp (enough) break after all, so that the third alternative, view *C*, can be maintained. **Clark** claims as much: We might "still discover that there is some functional property . . . which is both necessary for a content to become consciously known and yields an absolute temporal order of experiences. . . . I cannot see that this possibility is ruled out by anything that D & K tell us."

Suppose, then, that what makes some contentful brain events conscious is a property K that has a rather clear-cut onset. On such a view, contentful events, like plants, have rather long histories; they are unconsciously sown, develop, briefly bloom (rather suddenly acquiring some salient property K), and then fade into long-term memory or oblivion (losing property K). A single contentful event, let us suppose, can have a temporal subpart that is conscious, marked by the onset and offset of property K. To identify the subset of conscious events, just identify property K and motivate its identification. In order for any such claim to be taken seriously, some reason(s) must be given for singling out this property K, whatever it is, as the mark of consciousness (and hence the time of onset of K as the time of onset of consciousness). It will not do just to announce without further explanation that when events acquire property K, unlike their unconscious kin, they glow in the dark, as it were.

It is this independent motivation, we claim, that cannot be provided. There is no *further* functional or neurophysiological property K over and above the properties that account for the various "bindings" and effects on memory, speech, and other behavior, and *those* properties cannot distinguish between Orwellian and Stalinesque models. Once discrimination and control get distributed around to many subagencies, operating on different schedules, all the accomplishments of consciousness occur, one way or another, at one time or another, but no grounds remain for deeming one version of these events the "actual conscious experience."

The **EDITORIAL COMMENTARY** provides a minimal test case: a "simple, punctate, 'pink-elephant-now' thought." Surely *its* onset in consciousness is unproblematically clockable! As the Editor writes, "all we want to know is when the actually experienced *now* actually occurred." Let's consider the details. However "simple" and "punctate" it appears to be, the thought has to be *composed*, for it consists in (at least) three content components, *pink*, *elephant*, and – a tricky one – *now*, which must be generated and "bound" together somehow (the thought is of an elephant that is pink, and pink now, not just the thought of *elephant*, and *pink*, and *now*). The first two components may be generated in different regions, at slightly different times, and could presumably even acquire property K at different times. Or presumably they could get bound together but fail to acquire property K and fade away without ever being experienced, an *unconscious* pink-elephant-now content. This raises the question: Does the content *pink* become conscious before or after it is bound to *elephant*? That is, does the "binding" happen before property K is acquired by each element (the Stalinesque theory) or do the content elements acquire property K *seriatim*, perhaps in the order that they are generated in the brain, with the binding taking place only after all the elements are conscious (the Orwellian theory)? Or does the question not make sense (our

theory)? An analysis of the neural invariants recorded over hundreds of variations of pink-elephant-now thoughts may show precisely what conditions ensure that such a thought occurs at roughly such-and-such a time, and may even show considerable variation in time of binding relative to such other temporal landmarks as onset of *pink* activity and onset of motor response formation activity, but the one landmark that matters has no independent anchor: onset of *K*. The subject cannot provide any clues unobtainable by an analysis of the neural activity, because *ex hypothesi* the different instances are all the same subjectively – simple and punctate.

But what about "the actually experienced *now*"? The content *now* is tricky because, unlike *pink* and *elephant*, it is indexical: *It refers to its own vehicle of representation*, and hence bridges the gap between temporal properties represented and temporal properties of representings. But this fact creates a problem instead of solving one: The "concurrent neural event that corresponded to *now*" (the tokening of *now*, in philosophers' jargon) is presumably spread over a preconscious forming period and a postconscious remembered period (like a postcard on which you write "Please visit me *now*" and later mail to a friend, on whose desk it sits for weeks). Unless one can *independently* identify the onset of consciousness, the discovery of a "*now*" token in the brain hopelessly underspecifies a time of intended "utterance." The actually experienced *now* is presumably not the preparatory, forming-but-not-yet conscious *now*, or the reverberating-in-memory *now*, but the blooming-with-property-*K now*. How the Editor proposes to identify the thresholds that mark the transition of the tokening of *now* between these three states remains a mystery, because no functional difference has been shown to depend on these transitions and all the subjective evidence concerns only the temporal properties *represented* in the thought (e.g., the content: *"Now" is/was simultaneous with "pink-elephant"*), not the temporal properties of the representing. So even in the case of such a simple "punctate" thought, in which no questions arise about its veridicality relative to any external world it is portraying, when the Stalinesque/Orwellian distinction looms, nothing could settle it.

It is unclear to us why the Editor chose the epithet "Reaganesque" for his example, but it strikes us in any case as peculiarly apt for making *our* point, not his. Recall the question that rang through the land during the Iran-Contra hearings: What did the President know *and when did he know it*? This question mattered only because it was presumed that the President was actually in charge, actually a responsible decision-maker. But (according to widespread opinion) while the Great Communicator was an excellent public relations officer, he was otherwise a mere figurehead whose authority was more a convenient fiction than the wellspring of decision. If this is so, there were surely many times when it didn't make any political difference whether Reagan *actually* said the things that his staff had decided it would be good for him to say – *or to have said.* He could "fill in" the words if he wished, but if he didn't, no harm would be done. In the case of Reagan, there actually always was a (negligible) fact of the matter, but unless you think there is a Reagan homunculus in you, whose activities are for some reason constitutive of consciousness, there is no reason to believe in the analogous fact of the matter for "Reaganesque" events in your brain. In systems in which the authority and decision-making are distributed as they were in the Reagan White House and are in the brain, there is a sort of Virtual President but there is no fact of the matter about exactly where and when a Virtual President announces to himself "pink elephant now."

The onset of property *K* (whether it is gradual or sharp) is a temporal property of the representing, not a temporal property represented and, as all agree, we must be careful not to confuse these. There is a danger of confusion precisely because there is bound to be some sort of regular relation between them: It is no accident that representations of sequences of events will *tend* to be represented by sequences of representations of those events, that events represented *as* simultaneous will *tend* to be simultaneously represented, and so forth. A similar point applies to spatial properties: The representation of *x surrounded by a field of y* will *tend* to be accomplished by a representation of *x* surrounded by a representation of *y*, and so forth, but the space in the brain in which the representation is accomplished *is not* the space represented, however regular the correlation between spatial properties represented and spatial properties of the representation.

Rollins makes two points in this regard that we accept: (1) Temporal isomorphism may be required for some temporal tasks, and (2) the appeal to isomorphism need not commit what he calls the canonical fallacy. Sometimes the freedom to order events shrinks for all practical purposes to zero and isomorphism returns. It is easier to understand this point if we remind ourselves of its spatial analogue: Even though there is no general reason why the brain should always represent spatially contiguous regions of the world (or the body, or the retinal image of the world) by spatially contiguous vehicles of representation, there are circumstances in which the exigencies of engineering make this a very good, well nigh inevitable, solution to the design problem. But even when this is the case, we must be careful not to view the spatial isomorphism as a determining feature of the representational content rather than a limiting condition on the freedom of the vehicles. We must also not make the mistake of "restoring" the isomorphism in some dimly imagined later process. (The back-to-back semicircles of excitation that occur on the occipital cortex when one is visually stimulated by a large circular ring don't have to be reunited into a circular representation somewhere in order to be the brain's way of representing a circle. The same moral applies to time.)

3. When does "filling in" happen?

Shepard, while largely in agreement with us, claims we have gone too far in disparaging "filling in." We agree with Shepard that the target article permitted the inference to be drawn that the brain *never* bothers doing something like filling in and that this would be an overstatement. His work has shown clearly that there are phenomena in which there are analog or roughly continuous processes that do amount to a sort of filling in: As he puts it, these processes "successively activate (or 'fill in') intermediate states corresponding, in concrete detail, to what would be

successive intermediate states in the external world." **Roskies & Wood** make the same point citing further phenomena in which a sort of filling has been shown to occur. Their discussion also provides oft-repeated warnings – they appreciate the need for eternal vigilance – against the dangers of overinterpretation: "One cannot, of course, conclude that these particular neuronal effects are the only basis or the most important one for the subjective experience of color-filling." Indeed one must be careful not to jump to the conclusion that a particular instance or neuronal filling in (or, as the connectionists would say, vector completion) should be viewed as the basis, *as opposed to an effect*, of subjective experience. Vector completion – for example, in remembering – is not filling in of the sort we were criticizing. Vector completion is not "paint" used to render the *evidence* for some ulterior conclusion; vector completion *is* the conclusion. No more "giving" is needed because no more "taking" is going to occur. Such a process has the nice property of merging the distinction between presentation and judgment, turning the corner from given to taken without turning the corner sharply. Notice how we can accommodate Shepard's phenomena, for instance, without lapsing back into the Cartesian Theater. Suppose, in his illusory motion experiment, a visual subsystem in the subject's brain has arrived at the microjudgment that the vertical bar moved clockwise. But of course not *just* clockwise, but at a particular rate and during a particular window of (represented) time. Suppose this elaborated microjudgment is represented by an analog process, which in turn provides the grounds for a second microjudgment, provoked by the precisely timed probe spot flashed along the illusory trajectory of the moving bar, to the effect that the spot flashed before or after the bar had crossed that point. Each bit of judgment becomes the basis for further judgment, blurring the distinction between given and taken. The taking that there was motion happened before the taking that the spot was such and such – it had to, logically – but it also provided the basis for the judgment that in midcourse the dot was in such and such a location.

Such processes do provide, as **Roskies & Wood** claim, "a candidate explanation for observers' reported subjective experience," which *does not mean* that they provide an explanation of the construction of the "given" that is then subsequently taken by some further process of subjective experiencing. In their discussion of apparent motion, they propose a testable hypothesis:

> The perception of apparent motion might begin with an initially nondirectional priming from the first flashed stimulus. If another stimulus with similar characteristics should appear within the primed region before the activation has decayed, a *motion signal* [our emphasis] might be generated, either by potentiating the intervening circuit or by sending a signal that is interpreted by other areas as coding motion.

The point we would make – and we gather that **Shepard** and **Roskies & Wood** would agree – is that a *motion signal* is not to be confused with a *motion picture*, even when it is accomplished by a process (which is not *itself* accessible to consciousness) that generates intermediate states. (For more on "filling in" and evidence of analog or "roughly continuous" processes in perception, see Dennett 1991b).

4. Consciousness as a "system" with a function

One of philosophy's virtues is that it sometimes makes explicit the assumptions that are more covertly driving the imaginations of theorists in other disciplines, exposing them for assessment. A fine case is provided by the commentaries; the assumption that might be formulated in the following slogan:

> Every real thing must be in its own box.

According to **Antony**, the most charitable interpretation of our position is "eliminativism": "Consciousness does not exist." He eventually arrives at the correct reading of our view: "Conscious experiences are temporally located, but the time scale appropriate for the measurement of neurophysiological events (e.g., milliseconds) is too fine-grained to determine nonarbitary temporal boundaries for experiences." From this, however, he draws the ominous – and mistaken – conclusion that "conscious experiences have no role in the functional organization of the brain." (See **Velmans** for a similar conclusion, and also his *BBS* target article, 1991.) The conclusion we would draw instead is that conscious experiences have no role in the functional organization of the brain *in virtue of meeting some criterion of consciousness*. Becoming a conscious experience does not clearly endow an event with potencies it previously lacked. Antony says it follows from this that "conscious experiences cannot be identified with functional states or processes," but this is true only in the sense that conscious experiences cannot be identified *as a type* with any particular functional type. That is a far cry from "eliminativism." The events that constitute conscious experiences do play functional roles that can be timed down to the millisecond, but when or whether they play those roles is independent of when or whether they make the cut into the elite circle of conscious events.

The assumption that consciousness cannot be real unless there is a consciousness subsystem can be seen playing a slightly less visible role in the commentaries of **Teghtsoonian, Baars & Fehling, Velmans,** and **Farah.** (**Bridgeman** clearly expresses the case against the assumption.)

Teghtsoonian argues for a "central monitor" on the grounds that cross-modal comparison is a demonstrably real phenomenon. (Contrary to what he supposes, we do not claim to offer any arguments, new or old, against dualism. We assume that dualism is a dead horse, but that Cartesian materialism is alive and well, and worth combatting.) We agree with him that very many experiments, by S. S. Stevens and others, demonstrate conclusively that people can do cross-modal comparisons, which involve getting all the information into the same physical system. This "single monitor" is called the brain. Teghtsoonian takes the brain's capacity to serve all these comparative purposes as evidence for the existence of monitoring "systems" and then goes on to say that presumably similar monitoring systems exist for other attributes. Depending on how we understand "system" this is either trivially true, or entirely unlikely. People can compare the experience of falling in love with the experience of contracting a disease, or going into battle, but this gives no grounds for thinking that there is a monitor *system* (a "module" as some would say) specifically made for this

comparison task. We can construct comparison "machinery" on the fly, *ad lib.*, as we need it – thanks to the richly interconnected brain that provides the underlying hardware. The risk we run if we start diagramming these feats with the "boxological" flow charts of functionalism is that we will assume that the boundaries of the boxes we draw have a salience and integrity that project beyond the cases that inspired us to draw them.

(Our only misgivings about **Lycan**'s "inner sense" version of "UnCartesian materialism" concern his eagerness to reify all the episodes of corner-turning from given to taken into the acts of "internal scanners," which then encourages him to assert that "An internal monitor is an attention mechanism that can presumably be directed upon representational subsystems and stages of same." The difference in emphasis is slight, but the effects multiply. Before you know it, you are apt to have a warehouse of Crick-style (Crick 1984) searchlights illuminating – for their operators? – various attention-grabbing scenes. We urge Lycan to read **Bridgeman**'s concluding point: "Consciousness is not a monitor of mental life but a result of mental operations separated from the immediate sensorimotor world.")

In the same vein, our only quarrel with **Baars & Fehling** concerns their assumption that they cannot be realists about consciousness unless they consider their global workspace to be a "system" of limited capacity (and every item is either inside or outside the system), instead of saying that the brain has a limited capacity to perform the feats described on the left of their Table 1, and its performance of them can be seen as constituting the establishment of something rather like a global workspace (see also **Bridgeman,** who expresses the same view, which is elaborated in Dennett 1991b). Responding to **Reingold**'s well-put request, we would view Baars & Fehling's Table 1 not as marking distinct qualitative marks of consciousness but as perspicuously describing two poles (there are others) between which quantitative (but not easily quantified!) variation establishes the intuitive opposition between unconscious and conscious.

The difference between **Velmans**'s "integrationist" model and our Multiple Drafts model can be very succinctly stated: We *agree* with him that there are distributed processes of integration that tend, normally, in the fullness of time, to achieve a relatively integrated version of the world – the stream of consciousness that emerges in subjects' introspective and retrospective protocols; we just insist that there is no bridge across the stream! Velmans claims that in his view "within any given time window, information is integrated into a single conscious stream and that unless there is evidence to the contrary, subjects' reports about what they experience are accurate." If one were to expand this time window of consciousness to include every instant between initial stimulation and ultimate report, one would rule out all pre-experiential *and* post-experiential editing: For example, the subject would be held to be just as conscious of a raw retinal image as of any subsequent interpretations of it. If one tries to shrink the window so that all adjustments of content can be located as pre-experiential or post-experiential, one must draw in the bridge across the stream. What counts as "evidence to the contrary" depends on where one draws these boundaries (consider the insanely radical Orwellian hypothesis that we *are* conscious of the dizzying swim of imagery on our moving retinas in spite of what we subsequently say – we just continually forget it). Velmans notes, correctly, that the Orwellian view is in danger of being unfalsifiable. Indeed, and the same danger faces its Stalinesque twin.

Farah provides a nice demonstration of a familiar sort of theoretical blindness: Since you can (if you insist on it) redescribe *any* phenomena without mentioning consciousness, you can convince yourself that models of those phenomena that do mention consciousness tell us nothing about it. The target article is about consciousness because the phenomena in question are invisible to any research program that does not *interpret* subjects' protocols as accounts of conscious experience. If you treat the subjects' verbal behavior as so much emitted noise, the questions about subjective sequence, for instance, just cannot arise. Farah says that we "have not, so far, shed any light on the mechanisms of consciousness." What does she think these are, if not the very mechanisms of which we treat? If she means that there are *additional* mechanisms of consciousness, she has simply revealed her own unregenerate Cartesianism, materialist or dualist. If Farah is merely insisting that it is open to the theorist to claim that a subject who performs speech acts apparently descriptive of conscious experience might be a zombie just unconsciously mouthing the words, we would agree, but we did not think it necessary to direct our argument to those who still believe in zombies.

More explicit exploitation of the assumption that if it is real, it must be a system, can be found in the commentaries of the philosophers, **Lloyd, Block,** and **Van Gulick.** It beguiles **Lloyd** when he misattributes view *C* to us, claiming that our theory is not a theory of consciousness at all because "some parts of the multiple parallel streams are conscious, and some are not." Rather, we claim that there is no crisp way of telling exactly which parts of the multiple parallel streams are conscious. Any one of the streams sometimes contributes to awareness and sometimes not. No one stream is necessarily conscious by its very nature (see Kinsbourne, in preparation, for an extended discussion). And that *is* our theory of consciousness. It is a bad day for "realism" when only phenomena with hard-edged necessary and sufficient conditions can be considered real. More pointedly, when Lloyd goes on to declare that our "antirealist" arguments "collapse in the face of the reality of conscious experience, as undeniable for us as it was for Descartes," he pledges his allegiance to one of the most persistent illusions of Cartesian materialism: how it *seems to be* fixes how the *seeming must be.* Consciousness seems to be all or nothing, but as undeniable as *that* fact is, it does not license the conclusion that consciousness *is* all or nothing, that either there is a box in which consciousness, and nothing but consciousness, exists, or consciousness does not exist at all.

Block proposes that the seeming is accomplished in a "Cartesian Module," the one functional place in the brain that provides us with "phenomenal consciousness." He claims that we mistakenly assume without argument that there is no such thing as phenomenal consciousness. Well, what is it? Block supposes that we all just know: "Our fundamental access to phenomenal consciousness derives from our acquaintance with it." Whatever phenomenal consciousness is, Block tells us it is not any of the sorts of consciousness captured by such functional prop-

erties as "inferential promiscuity" (essentially, the property represented on the left of Table 1 in **Baars & Fehling**) or accompanying second-order thoughts (see **Rosenthal**). Block sees an opening for what he thinks will be a nonarbitrary way of sorting brain events into the conscious and unconscious: "What if some of the brain representations of an event . . . are phenomenally conscious whereas others are not?" Indeed, what if? Shall we find out which these are by asking subjects? If so, all the arguments of the target article apply, for the quandaries arise precisely because of the problems encountered in interpreting the protocols or introspective claims of subjects. Does Block then hold, perhaps, that it is only possible to determine correlations of phenomenal consciousness with brain events by autocerebroscopy? Even here we will run into problems, according to Block's own account, for even the most careful and persistent autocerebroscopist is stuck relying on his own judgments of what seems to be happening, but "to assume that the subjective is exhausted by judgments is to beg the question against phenomenal consciousness, which, if it exists, is *not* just a matter of judgments." So a subject studying himself with an autocerebroscope (and his native introspective talents) will not be able to avoid begging the questions against the existence of phenomenal consciousness.

It may be folly for us to attempt to say anything more about a putative phenomenon so sublimely inaccessible to investigation as *phenomenal* consciousness, but since both **Block** and **Van Gulick** venture to propose models of it, a comment is in order. The main attraction of Stalinesque theories, apparently, is that they agree with Descartes that the Show Must Go On, and it is worth noting that both models are Stalinesque. This is explicitly stated in the case of Van Gulick's GIPS model. It is less obvious in the case of Block's Cartesian Modularism, the view that "all conscious events occur in a single system," but it follows from the supposition that events occurring in a single system will be the effects of events occurring in other systems – the idea is that *after* some unconscious discriminatory machinery does its work, it sends some effect to the Consciousness Module so that its results can be properly bathed in phenomenal properties. We are in no position to assert that *any* model of "phenomenal" consciousness would have to be Stalinesque, but it looks like a good bet.

Consider **Van Gulick**'s GIPS model. Like **Block**, he claims that his model "draws a clear distinction between phenomenal and nonphenomenal representation," but he similarly does not venture any account of what this distinction is. He claims that neurological evidence could be used to secure it, however, and in his account of metacontrast supposes that we could confirm or disconfirm his Stalinesque interpretation by relying on "what we *independently* [our emphasis] know to be the neural correlate of phenomenal representation of the relevant stimulus." But this cannot just be asserted – as if we could tell by looking at the neurological evidence that some of the neural events glowed in the dark. Our point is that when we turn to the only imaginable source of "independent" evidence for the distinction – the judgments of subjects, as expressed in their protocols (or our own judgments, if we use the autocerebroscope) – we find the Janus-faced evidence of Orwell/Stalin. Moreover, when we acknowledge, as Van Gulick insists, that such judgments cannot be viewed as infallible (he mistakenly supposes that our attack on the "straw man" of Cartesian materialism depends on this), we must admit the need for some Archimedean point that can impeach a particular judgment *on neurological grounds*, throwing us back on the requirement of motivating, independently, the identification of some neurological feature (some property *K*) as the mark of "phenomenal" representation. Can this be done? The target article claims that it cannot, drawing for support on a general argument that makes no specific mention of "phenomenal" consciousness. Block has given us no reason to think that phenomenal consciousness (whatever that is) would be more rather than less readily anchored to independent neural correlates, and as he notes, there are plenty of detailed arguments directed against the concept of phenomenal consciousness in Dennett (1991b).

5. Libet and Treisman tilt at philosophers

Libet and **Treisman** both reveal that they expect no illumination of the issues from the efforts of philosophers, and perhaps this shared attitude – and its accompanying low estimate of how informed they presume the authors to be – contributes to their egregious misreadings of the target article.

Libet makes it clear that he is not a dualist. We are happy to have given him the opportunity to disengage himself more sharply than before from his dualist supporters. We also accept his clarification of what we had viewed as an equivocation between a radical and mild version of "backwards referral." We note, however, that in the very sentence in which he claims to have been clear about the distinction, he uses a dangerously equivocal phrase; he claims to have demonstrated "the subjective timing of sensory experience as *appearing* [our emphasis] before the time of the adequate neural representation for the experience." In the absence of his declaration, one might well be tempted to read this the way the dualists did: as the claim that a sensory experience *actually* appears (blooms in consciousness) before the time of the adequate neural representation, rather than what he apparently means: It only *appears to appear* in consciousness before that time.

Libet makes it clear at the outset that his primary motive for preferring a Stalinesque model is that it seems to him to explain "the actual experience of a single narrative" – something our view, in his opinion, cannot handle. (**Aronson et al.** and **Van Gulick** also seem to think that it is easier to explain the observed unity of consciousness by positing a single Stalinesque theater, but as **Hurley** notes, "the unity of the collection of vehicles in the theatre provides only an illusion of understanding the unity of consciousness.") Often Libet simply *assumes* the Stalinesque interpretation of his experiments. When he does address the Orwellian alternative (which is not, as he thinks, *our* model), he misunderstands it. For instance, he thinks he can eliminate the Orwellian alternative to his Stalinesque account of metacontrast by claiming that Orwellian "intensification" of memory is more ad hoc than "obliteration," ignoring the fact that the Orwellian alternative was *introduced* by us as one in which memory "improvements" were inserted. And in his description of

his study of somatosensory thalamus stimulation (Libet et al. 1991) he mixes the Stalinesque and Orwellian interpretations of his results without noticing it: "Clearly, subjects had a memory of a brief stimulus, which they later correctly identified [in a forced choice guess], but no *reportable* [our emphasis] awareness." That is just what the Orwellian would say, insisting that this showed that subjects *were* (briefly) conscious of the stimulus in spite of the fact that it was not subsequently reportable.

Libet's response to the technical criticisms we mentioned is misplaced. We chose to direct attention to his interesting work in spite of the widespread opinion that they "involved very few subjects, were inadequately controlled, and have not been replicated." Referees and readers of early drafts of the target article had objected to our choice of his phenomena on these grounds (see also **Treisman**), so we felt obliged to acknowledge that opinion. In fact we share it, and are quite qualified, as is Churchland (see especially 1981b), to express criticisms about both his methods and interpretations. We do view the target article as supporting his conviction that there should be further attempts to replicate this difficult work.

Libet points out that there were three factors in addition to the one we cited for his estimate of the time of "neuronal adequacy" for consciousness, but the other three are more obviously question-begging in the context of fending off Orwellian alternative models. His discussion in sections 3.2 and 3.3 repeats his misunderstandings of the Orwellian alternative (and the idea that it is *our* alternative) mentioned above. For instance, he fails to notice that (1) the Taylor and McCloskey (1990) results are neutral between an Orwellian and Stalinesque treatment, (2) his claim that "the experimental evidence is precisely that the initiating neural events produce conscious intention, but only after a delay of about 350 msec" is itself question-begging, and (3) in his analysis of the Grey Walter experiment, he is not entitled to take the subjects' reports at face value as an indication of the order in which the crucial events *actually* happened in the brain.

Treisman misreads the target article, and then directs his lengthy discussion against a fantasy of his own construction. Most of his points are simply versions of points we make ourselves, or thought too obvious to need rehearsal, and his assertions about the implications of our view are almost all far wide of the mark. Perhaps we are to blame, but others did not misread it as radically as he appears to have done. For instance, Treisman says there is no paradox; that was our point. The paradox arises only when one interprets the events through a mistaken view, the Cartesian Theater. He reads the questions we put in the mouth of a naive mechanist as questions we ourselves set out to answer, when in fact we set out to discredit those questions much as he does (we are not Orwellians, a fact that has escaped him).

He claims that we make two major errors. First, we confuse two different senses of "represent." He cites no instances, and in fact we use the term always in a sense that carries a connotation of content that is potentially usable by the system. Second, we suppose that "subjective sequence is "determined in consciousness." This idea that we "fail to distinguish between neural processing on the one hand and hypothetical processes proceeding 'in consciousness' on the other" is a wild misreading, since it is precisely our point that there *is no such arena* in which anything could be determined by any hypothetical processes. *Of course* it is neural processing that determines subjective sequence. Our point is that once this has happened, there is no *further* phenomenon: There is no "running in proper sequence" or "projection" of the contents whose order has been fixed by those neural processes.

Once **Treisman**'s views are sorted out, they emerge, not as he thinks, as an alternative to ours, but as a truncated or underspecified version of ours. For example, he agrees with us that thanks to the invariance of what he calls relation *U* the subjective sequence need not mirror the objective sequence of events in the brain. And he also says that "we can safely conclude that the extent of the delay before the neural mechanisms reach a perceptual decision on an input may vary with the context in which the subject operates and the nature of the task." Of course. But he does not go on to address the question of whether this implies that consciousness of the relevant events is in all respects delayed until all such perceptual decisions have been made. Does Treisman think, for instance, that there is no consciousness at all of the second tap in a cutaneous rabbit test until the brain has had a chance to arrive at the perceptual decision that the tap is displaced? Until he actually confronts such questions, we won't know whether Treisman is in complete agreement with us, or is seduced by either an Orwellian or a Stalinesque vision.

6. Other objections

Wasserman (see also **Libet**) points out that the problems of temporal smear – ranging up to several hundred milliseconds – are not due *directly*, as we suggested, to the slow speed of axonal *transmission* but rather to the *competition* between larger patterns of neuronal activity that take time to be resolved. The fact remains, however, that the speed limit for such resolution is ultimately a function of the basic cycle times for individual neural elements. Although we also agree with Wasserman that the ultimate basis for any neural integration must be the summation of graded potentials inside single cells, single neurons are too "stupid" to be responsible, on their own, for whole microtakings, so we must advance to the level of larger patterns of activity to find appropriate machinery for accomplishing such judgments.

We ventured no details on how to model these processes, but **Gregson** provides an elegant account of hypothesized dynamics of neural systems that predicts the existence in principle of temporal anomalies of consciousness; his analysis may contribute to an explanation of how the multiple draft machinery works. In particular, it provides a response to **Wasserman** (who argues that integration in the last analysis must take place by summation, on the minitheatrical stage of the individual neuron) by suggesting coupled neural dynamics as an integrative principle. Incidentally, Wasserman's point that three or more coupled processes sometimes become chaotic appears neatly applicable to Norman's (1967) observation that when subjects rapidly identify successive sequences of visual or auditory stimuli, they often believe themselves to have experienced the last two stimuli in reverse order. Clearly, according to Gregson, such an occurrence

does not necessitate belief in a sequence concretely located in the brain.

Wasserman usefully elaborates the details of the Kolers and von Grünau (1976) experiment with color phi, but we would hold that he is mistaken in claiming that their choice of dependent variable, having the subjects do a matching task instead of giving an oral response, altered the situation in such a way as to avoid the Orwell/Stalin quandary. Given the instructions, when a subject leaves the knob in a particular orientation, this amounts to uttering a predefined speech act, a more refined and accurate speech act than would be verbally achievable, but still the product of judgmental processes that allow room for the quandary to arise.

Warren is in more agreement with us than he realizes. He slightly misreads our remark about how, as a matter of logic, certain discriminations must amount to temporal discriminations. We agree with him that "patterns are recognized first, and then properties of these patterns inferred" – when they are, which is seldom the case for some such properties. Consider the same point in a different domain. If subjects can distinguish between tachistoscopic presentations of the words "back" and "buck," this must be, as a matter of logic, a case of being able to distinguish two stimuli that differ only in one property: The presence or absence of the tiny bit of black that distinguishes the "*A*" from the "*U*" in the typeface used. But of course the subject doesn't make this discrimination by first looking for that bit of black and then drawing an inference about the identity of the word.

Block begs to differ from our interpretation of the quotations of Harnad (1982; 1989), Mellor (1981), and Churchland (1981b), and invites the reader to look again. We second the invitation. In his discussion of the Mellor quote, he says "the issue here is not one of judgments, but of experience of relations among experiences." We take **Van Gulick** to attempt the same defense of Mellor. This seems to mean that the experience of having the experience of *A*-as-happening-before-*B* must be the experience of having the experience of *A* before having the experience of *B*, but Mellor offers no argument for this (deeming it self-evident, we presume), and as we argued in the target article, there are reasons for abandoning this traditional prejudice. **Glymour et al.** produce a passage that succinctly illustrates the fundamentally tempting confusion: "Suppose a temporal sequence composed of experiences *E*1, *E*2, *E*3, in that order. Either *the content is experienced in that order too* [our emphasis], or that temporal sequence is not experienced at all." To sustain their claim, the italicized phrase *must* be read to carry no implications at all about whether the content is experienced *as* in that order, but this is hardly a familiar reading. Suppose, for instance, that the temporal sequence in question is a dream and suppose it is actually dreamt backwards (*E*1 is the shocking conclusion, *E*2 is the midgame, *E*3 is the opening scene). Now if dreams are experiences, then either the content is experienced in the order conclusion-midgame-opening or it is not experienced at all, but it must not be supposed to follow that such a dream would be experienced as anomalous in any way (Dennett 1976). If the dreamer found the whole experience anomalous, this would not be because of the order in which the content was experienced but because, for one reason or another, the content was experienced *as being* in the anomalous order.

Shepard faults us for claiming that no one draft is more "correct than another." Here he overlooks our intended meaning, which was not, as he supposes, that no one draft is apt to be a more veridical account of what happened in the environment and in the control structure of the agent (we agree with his remarks about this), but simply that no subsequent introspective report can claim to be a veridical account of "what happened in consciousness" – as opposed to an account of what happened in the brain one way or another.

Aronson et al. offer figures that do not accurately represent the claims made in the target article. Their Figure 1a shows Orwellian editing to be preconscious, and Figure 1b apparently models view *C*, not the multiple drafts view (the same misreading found in **Lloyd**). We are unclear about how to interpret their preferred diagram 1c, but the lines from "local discriminations" to the "functional theater" make it appear to be some version of the Cartesian Theater after all.

7. Supporting arguments

We want to express our appreciation to **Damasio, Hurley, McDermott, Reingold, Rosenthal,** and **Young** for commentaries that support our position with considerations drawn from very different quarters.

Damasio, Reingold, and **Young** all confirm our conviction that, in spite of disclaimers, allegiance to the imagery of the Cartesian Theater is a widespread bad habit of thought, not a straw man as some commentaries suggest.

Damasio cites neuroanatomical grounds for resisting this temptation, which is a valuable antidote (though not, of course, a refutation) to **Libet**'s claim that "the existence of such a single locus is not yet experimentally excluded." Damasio says, correctly, that a satisfactory model of consciousness should indicate how the unintegrated fragments operate to produce the integrated self; we have tried to respond to this demand in sections 1 and 2 above. We agree that somatic states play an ineliminable role in any robust integration of consciousness, but the issues raised by the specifics of such a proposal are beyond the scope of our discussion.

Damasio also alludes to Sherrington's idea, recently refurbished by von der Malsburg (1987) and others, that – one might say – timing is binding. We agree that there may well be something deeply right about this idea but if so, two implications should be noted: (1) The timing that binds is not a photo-finish but a phasic relation between concurrent events whose neural consequences have a considerable temporal spread, and (2) insofar as simultaneity of phase behavior plays the role of pure binding, it necessarily becomes unavailable to play the role of a basis for a judgment of simultaneity. (For example, the successive notes of a melody are bound but certainly are not *represented as* occurring simultaneously, nor does the use of delay lines to create a simultaneous representation of the successive notes appear to be a workable mechanism for binding. For an analysis of some of the relevant issues, see Port 1990, and Port & Anderson 1990.)

Young uses the McGurk effect to support the Multiple Drafts model by showing how *un*reliable introspection is

as a source of knowledge about how what seems to happen actually happens.

Rosenthal and **Hurley** both note that a particularly tempting traditional Cartesian assumption is that whatever consciousness is, it is an intrinsic property of those mental states that are conscious. Both then show, by somewhat different routes, how this assumption suffices to make the Cartesian Theater seem inevitable. Giving up consciousness as an intrinsic property, Rosenthal recommends that we adopt the term *transitive consciousness* for what we have called property *K*: the property the acquisition of which makes a mental state conscious. At first this appears to open up what might be called a forced choice for the theorist: Orwellian or Stalinesque? But he shows that any answer will be, as we claim, arbitrary. Hurley arrives at the same verdict. These commentaries provide valuable support for our position, by showing how its most "radical" conclusions are consistent with a traditional metaphysical position that relinquishes only one dubious feature of the Cartesian heritage: the idea of consciousness as an intrinsic property.

Hurley also anticipates and rebuts the objection that our view is "verificationist," and in particular notes that "to object that the 'observer' to whom manifestation is being required is oneself at a slightly later time is to assume the temporal atomism about consciousness that is [our] target."

McDermott speculates about the underlying moral anxieties that may well provide a hidden agenda in support of Cartesian materialism. As he says, in a forthright rejection of this theme, "if people are valuable, it is not because they are imperishable souls connected to bodies only for a brief sojourn." There is more on this important topic in Dennett (1991b).

References

Letters *a* and *r* appearing before authors' initials refer to target article and response respectively.

Allport, A. (1988) What concept of consciousness? In: *Consciousness in contemporary science*, ed. A. J. Marcel & E. Bisiach. Cambridge University Press (Cambridge). [aDCD]

Armstrong, D. M. (1968) *A materialist theory of the mind.* Routledge & Kegan Paul. [NB, WGL]

Baars, B. J. (1983) Consciousness provides the nervous system with coherent, globally distributed information. In: *Consciousness and self-regulation*, vol. III, ed. R. Davidson, G. Schwartz & D. Shapiro. Plenum Press. [BJB]

(1988) *A cognitive theory of consciousness.* Cambridge University Press (Cambridge). [BJB, MV]

Barlow, H. B. & Levick, W. R. (1965) Mechanisms of pattern selectivity in retina. *Journal of Physiology* 178:477. [aDCD]

Bergson, H. (1910) *Time and free will.* Harper & Row. [MJF]

Bisiach, E. (1988) The haunted brain and consciousness. In: *Consciousness in contemporary science*, ed. A. J. Marcel & E. Bisiach. Cambridge University Press (Cambridge). [aDCD]

Block, N. (1978) Troubles with functionalism. In: *Mind and cognition* ed. W. Lycan. Blackwell. [NB]

(1991) Evidence against epiphenomenalism. *Behavioral and Brain Sciences* 14(4):670. [NB]

(forthcoming) Does consciousness have a function? [NB]

Blumenthal, A. L. (1977) *The process of cognition.* Prentice-Hall. [MV]

Boring, E. G. (1950) *A history of experimental psychology*, 2nd ed. Appleton-Century-Croft. [MT]

Breitmeyer, B. G. (1984) *Visual masking.* Clarendon Press. [aDCD]

(1985) Problems with the psychophysics of intention. *Behavioral and Brain Sciences* 8:539–40. [aDCD]

Bridgeman, B. (1985) Free will and the functions of consciousness. *Behavioral and Brain Sciences* 8(2):540. [aDCD]

(1988) *The biology of behavior and mind.* John Wiley. [BB]

(1990) Empirical work on conscious vs. unconscious processes. *Report #54/1990, Research group on Mind and Brain*, Zentrum fuer interdisciplinaer Forschung, University of Bielefeld, Germany. [BB]

Calvin, W. (1990) *The cerebral symphony: Seashore reflections on the structure of consciousness.* Bantam. [aDCD]

Campbell, R. (1989) Lipreading. In: *Handbook of research on face processing*, ed. A. W. Young & H. D. Ellis. North Holland. [AWY]

Castiello, U., Paulignan, Y. & Jeannerod, M. (1991) Temporal dissociation of motor responses and subjective awareness. A study in normal subjects. *Brain* 114. [MJ]

Cheesman, J. & Merikle, P. M. (1986) Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology* 40:343–67. [EMR]

Churchland, P. S. (1981a) On the alleged backwards referral of experiences and its relevance to the mind-body problem. *Philosophy of Science* 48:165–81. [aDCD, NB, BL, ALR]

(1981b) The timing of sensations: Reply to Libet. *Philosophy of Science* 48:492–97. [arDCD, NB]

Cooper, L. A. (1976) Demonstrations of a mental analog of an external rotation. *Perception and Psychophysics* 19:296–302. [RNS]

Cooper, L. A. & Shepard, R. N. (1973) Chronometric studies of the rotation of mental images. In: *Visual information processes*, ed. W. G. Chase. Academic Press. [RNS]

Corkin, S. (1984) Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings. In: *Seminars in Neurology* 4:249–59. [ARD]

Crick, F. (1984) Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Science* 81:4586–90. [rDCD, ARD]

Crick, F. & Koch, C. (1990) Towards a neurobiological theory of consciousness. In: *Seminars in the neurosciences*, 2:263–75, ed. A. R. Damasio. W. B. Saunders. [aDCD, NB, ARD]

Damasio, A. R. (1989a) Time-locked multiregional retroactivation: A systems level proposal for the neural substrates of recall and recognition. *Cognition* 33:25–62. [aDCD, ARD]

(1989b) The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation* 1:123–32. [ARD]

(1990) Synchronous activation in multiple cortical regions: A mechanism for recall. *Seminars in the neurosciences* (The "Neurobiology of Mind" issue) 2:287–96. [ARD]

Damasio, A. R., Tranel, D. & Damasio, H. (1989) Amnesia caused by herpes simplex encephalitis, infarctions in basal forebrain, Alzheimer's disease, and anoxia. In: *Handbook of neuropsychology*, ed. F. Boller & J. Grafman, vol. 3, ed. L. Squire. Elsevier. [ARD]

Danto, A. (1985) Consciousness and motor control. *Behavioral and Brain Sciences* 8:540–41. [aDCD]

Deecke, L., Grozinger, B. & Kornhuber, H. H. (1976) Voluntary finger movement in man: Cerebral potentials and theory. *Biological Cybernetics* 23:99–119. [BL]

De Haan, E. H. F., Young, A. & Newcombe, F. (1987) Face recognition without awareness. *Cognitive Neuropsychology* 4:385–415. [MJF]

Dennett, D. C. (1976) Are dreams experiences? *Philosophical Review* 85:151–71. [rDCD]

(1978) Skinner skinned. In: *Brainstorms: Philosophical essays on mind and psychology*, ed. D. C. Dennett. Bradford Books. [aDCD]

(1979) [review of K. R. Popper and J. C. Eccles 1977.] *Journal of Philosophy* 76:91–97. [aDCD]

(1982) How to study human consciousness empirically. *Synthese* 53:159–80. [aDCD]

(1988) Quining Qualia. In: *Consciousness in contemporary science*, ed. A. J. Marcel & E. Bisiach. Cambridge University Press (Cambridge). [NB]

(1991a) Real patterns. *Journal of Philosophy* 87:27–51. [rDCD]

(1991b) *Consciousness explained.* Little Brown. [arDCD, NB]

Descartes, R. (1662) *Traite de l'homme.* Paris. [aDCD]

(1662/1972) *Treatise on man.* Translated by T. S. Hall. Harvard University Press. [MV]

Dietrich, E. (1985) Computer thought: Propositional attitudes and metaknowledge. Ph.D. dissertation, Department of Philosophy, University of Arizona. [JA]

Dimond, S. J. (1980) *Neuropsychology: A textbook of systems and psychological functions of the human brain.* Butterworths. [MV]

Dixon, N. F. (1971) *Subliminal perception: The nature of a controversy.* McGraw-Hill. [EMR]

(1981) *Preconscious processing.* John Wiley. [EMR, MV]

(1986) On private events and brain events. *Behavioral and Brain Sciences* 9:29–30. [aDCD]
Dorfman, A. (1988) *Mascara*. Viking. [aDCD]
Dorman, M. F., Cutting, J. E. & Raphael, L. J. (1975) Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance* 104:121–29. [RMW]
Dunn, J. D. & Kirsner, K. (1988) Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review* 95:91–101. [EMR]
(1989) Implicit memory: Task or process? In: *Implicit memory: Theoretical issues*, ed. S. Lewandowsky, J. C. Dunn & K. Kirsner. Erlbaum. [EMR]
Eccles, J. C. (1980) *The human psyche*. Springer. [MV]
(1985) Mental summation: The timing of voluntary intentions by cortical activity. *Behavioral and Brain Sciences* 8:542–43. [aDCD]
Edelman, G. (1989) *The remembered present*. Basic Books. [ARD]
Efron, R. (1967) The duration of the present. *Proceedings of the New York Academy of Science* 138:713–29. [aDCD]
(1973) Conservation of temporal information by perceptual systems. *Perception and Psychophysics* 14:518–30. [aDCD]
Englekamp, J. (1988) Modality-specific encoding and word class in verbal learning. In: *Practical aspects of memory: Current research and issues*, vol. I, ed. M. M. Gruneberg, P. E. Morris & R. N. Sykes. John Wiley. [MR]
Erdelyi, M. H. (1985) *Psychoanalysis: Freud's cognitive psychology*. Freeman. [EMR]
(1986) Experimental indeterminacies in the dissociation paradigm of subliminal perception. *Behavioral and Brain Sciences* 9(1):30–31. [aDCD, EMR]
Eriksen, C. W. (1960) Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review* 67:279–300. [EMR]
Farah, M. J., Monheit, M. A. & Wallace, M. A. (1991) Unconscious perception of extinguished visual stimuli: Reassessing the evidence. *Neuropsychologia* 29:949–58. [MJF]
Farrell, B. A. (1950) Experience. *Mind* 59:170–98. [aDCD]
Fehling, M. R., Altman, A. M. & Wilber, B. M. (1989) The heuristic control virtual machine: An instance of the Schemer computational model of reflective real-time problem-solving. In: *Blackboard systems and their applications*, ed. V. Jagganathan, R. Dodhiawala & L. Baum. Academic Press. [BJB]
Fehling, M. R., Baars, B. J. & Fisher, C. (1990) A functional role for repression in an autonomous, resource-constrained agent. *Proceedings of the XII Cognitive Science Conference*, Boston, MA. [BJB]
Felsten, G. & Wasserman, G. S. (1980) Visual masking: Mechanisms and theories. *Psychological Bulletin* 88:329–53. [GSW]
Flanagan, O. (1991) Consciousness. In: *The science of the mind*, 2nd ed. MIT Press. [NB]
Gallistel, C. R. (1990) *The organization of learning*. MIT Press. [aDCD]
Geldard, F. A. (1977) Cutaneous stimuli, vibratory and saltatory. *Journal of Investigative Dermatology* 69:83–87. [aDCD]
Geldard, F. A. & Sherrick, C. E. (1972) The cutaneous "rabbit": A perceptual illusion. *Science* 178:178–79. [aDCD, MT]
(1983) The cutaneous saltatory area and its presumed neural base. *Perception and Psychophysics* 33:299–304. [aDCD]
(1986) Space, time and touch. *Scientific American* 254:90–95. [aDCD]
Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B. & Massey, J. T. (1989) Mental rotation of the neuronal population vector. *Science* 243:234–36. [RNS]
Gide, A. (1948) *Journal des faux monnayeurs*. Gallimard. [aDCD]
Glynn, I. M. (1990) Consciousness and time. *Nature* 348:477–79. [aDCD, BL]
Goodman, N. (1978) *Ways of worldmaking*. Harvester. [aDCD]
Gray, C. M., Konig, P., Engle, A. K. & Singer, W. (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338:334–37. [ARD]
Gregson, R. A. M. (1988) *Nonlinear psychophysical dynamics*. Erlbaum. [RAMG]
(1991) *n-Dimensional nonlinear psychophysics*. Erlbaum. [RAMG]
Grey Walter, W. (1963) Presentation to the Ostler Society, Oxford University, Oxford, England. [aDCD]
Haken, H., ed. (1985) *Complex systems – operational approaches*. Springer-Verlag. [RAMG]
Harnad, S. (1982) Consciousness: An afterthought. *Cognition and Brain Theory* 5:29–47. [arDCD, NB]
(1989) Editorial commentary [on Libet 1985a]. *Behavioral and Brain Sciences* 12:183. [arDCD]
(unpublished) Conscious events cannot be localized in time. [aDCD]
Hawking, S. (1988) *A brief history of time*. Bantam. [aDCD]
Hildreth, E. C. & Koch, C. (1987) The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience* 10:477–533. [ALR]
Hoffman, R. E. & Kravitz, R. E. (1987) Feedforward action regulation and the experience of will. *Behavioral and Brain Sciences* 10:782–83. [aDCD]
Holender, D. (1986) Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences* 9:1–23. [aDCD, EMR]
Honderich, T. (1984) The time of a conscious sensory experience and mind-brain theories. *Journal of Theoretical Biology* 110:115–29. [aDCD]
Hoy, R. C. (1982) Ambiguities in the subjective timing of experiences debate. *Philosophy of Science* 49:254–62. [aDCD]
Jacoby, L. L. & Whitehouse, K. (1989) An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General* 118:126–35. [EMR]
Jacoby, L. L., Woloshyn, V. & Kelley, C. (1989) Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General* 118:115–25. [EMR]
James, W. (1890) *The principles of psychology*, vol. I. Henry Holt. [MT]
Jasper, H. H. (1985) Brain mechanisms of conscious experience and voluntary action. *Behavioral and Brain Sciences* 8(4):543–44. [aDCD]
Javel, E., McGee, J. A., Horst, J. W. & Farley, G. R. (1988) Temporal mechanisms in auditory stimulus coding. In: *Auditory function: Neurobiological bases of hearing*, ed. G. M. Edelman, W. E. Gall & W. M. Cowan. John Wiley. [ALR]
Jeannerod, M. (1990) Traitement conscient et inconscient de l'information perceptive. *Revue Internationale de Psychopathologie* 1:13–34. [MJ]
John, E. R. (1976) A model of consciousness. In: *Consciousness and self-regulation*, ed. G. Schwartz & D. Shapiro. Plenum. [MV]
Julesz, B. (1971) *Foundations of cyclopean perception*. University of Chicago Press. [aDCD]
Jung, R. (1985) Voluntary intention and conscious selection in complex learned action. *Behavioral and Brain Sciences* 8(4):544–45. [aDCD]
Keller, I. & Heckhausen, H. (1990) Readiness potentials preceding spontaneous motor acts: Voluntary vs. involuntary control. *Electroencephalography and Clinical Neurophysiology* 76:351–61. [BL]
Kinsbourne, M. (1985) Parallel processing explains modular informational encapsulation. *Behavioral and Brain Sciences* 8(1):23. [rDCD]
(1988) Integrated field theory of consciousness. In: *Consciousness in contemporary science*, ed. A. J. Marcel & E. Bisiach. Oxford University Press (Oxford). [arDCD]
(in preparation) The distributed brain basis of consciousness. [rDCD]
Koch, C. & Crick, F. (1991) Understanding awareness at the neuronal level. *Behavioral and Brain Sciences* 14(4):683–85. [MV]
Kolers, P. A. (1972) *Aspects of motion perception*. Pergamon Press. [aDCD]
Kolers, P. A. & von Grünau, M. (1976) Shape and color in apparent motion. *Vision Research* 16:329–35. [arDCD, MT, GSW]
Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E. & Carr, C. E. (1988) Neurophysiological and anatomical substrates of sound localization in the owl. In: *Auditory function: Neurobiological bases of hearing*, ed. G. M. Edelman, W. E. Gall & W. M. Cowan. John Wiley. [ALR]
Kubovy, M. (1983) Mental imagery majestically transforming cognitive psychology. Review of *Mental images and their transformations*. *Contemporary Psychology* 28:661–63. [RNS]
Kulli, J. & Koch, C. (1991) Does anesthesia cause loss of consciousness? *Trends in Neuroscience* 14:6–10. [BG]
Lashley, K. S. (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior: The Hixon symposium*, ed. L. A. Jeffress. John Wiley. [MT]
Latto, R. (1985) Consciousness as an experimental variable: Problems of definition, practice, and interpretation. *Behavioral and Brain Sciences* 8:545–46. [aDCD, BL]
Levine, J. (1983) Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*. [NB]
Libermann, A. M. (1970) The grammar of speech and language. *Cognitive Psychology* 1:301–23. [aDCD]
Libet, B. (1965) Cortical activation in conscious and unconscious experience. *Perspectives in Biology and Medicine* 9:77–86. [aDCD, BL]
(1973) Electrical stimulation of cortex in human subjects, and conscious sensory aspects. In: *Handbook of sensory physiology*, ed. A. Iggo. Springer. [BL]
(1978) Neuronal vs. subjective timing for a conscious sensory experience. In: *Cerebral correlates of conscious experience*, ed. P. A. Buser & A. Rougeul-Buser. Elsevier/North Holland Biomedical Press. [BL]

(1981) The experimental evidence for subjective referral of a sensory experience backwards in time: Reply to P. S. Churchland. *Philosophy of Science* 48:182–97. [aDCD, BL]
(1982) Brain stimulation in the study of neuronal functions for conscious sensory experiences. *Human Neurobiology* 1:235–42. [aDCD, BL]
(1985a) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8:529–66. [aDCD, BL, EMR, RNS, MT]
(1985b) Subjective antedating of a sensory experience and mind-brain theories: Reply to Honderich (1984). *Journal of Theoretical Biology* 114:563–70. [aDCD, MT]
(1987) Are the mental experiences of will and self-control significant for the performance of a voluntary act? *Behavioral and Brain Sciences* 10:783–86. [aDCD, EMR]
(1989a) The timing of a subjective experience. *Behavioral and Brain Sciences* 12:183–85. [aDCD, BL, EMR]
(1989b) Conscious subjective experience vs. unconscious mental functions: A theory of the cerebral processes involved. In: *Models of brain function*, ed. R. M. J. Cotterill. Cambridge University Press. [BL]
(1991a) Conscious and neural timings. *Nature* 352:27. [BL]
(1991b) Voluntary acts and readiness-potentials. *Electroencephalography and Clinical Neurophysiology*. (in press) [BL]
Libet, B., Alberts, W. W., Wright, E. W., Delattre, L. D., Levin, G. & Feinstein, B. (1964) Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. *Journal of Neurophysiology* 27:546–78. [BL]
Libet, B., Alberts, W. W., Wright, E. W. & Feinstein, D. (1967) Responses of human somatosensory cortex to stimuli below threshold for conscious sensation. *Science* 158:1597–1600. [BL]
Libet, B., Gleason, C. A., Wright, E. W., Jr. & Pearl, D. K. (1983a) Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential); the unconscious initiation of a freely voluntary act. *Brain* 106:623–42. [BL]
Libet, B., Pearl, D. K., Morledge, D. M., Gleason, C. A., Hosobuchi, Y. & Barbaro, N. M. (1991) Control of the transition from sensory detection to sensory awareness in man by the duration of thalamic stimulus. *Brain* 114:1731–57. [rDCD, BL]
Libet, B., Wright, E. W., Feinstein, B. & Pearl, D. K. (1979) Subjective referral of the timing for a conscious sensory experience: A functional role for the somatosensory specific projection system in man. *Brain* 102:193–224. [aDCD, BL, RNS, MT]
Libet, B., Wright, E. W., Jr. & Gleason, C. A. (1983b) Preparation-or-intention-to-act, in relation to pre-event potentials recorded at the vertex. *Electroencephalography and Clinical Neuropsychology* 56:367–72. [BL]
Lloyd, D. (1989) *Simple minds*. MIT Press. [DL]
(1991) Leaping to conclusions: Connectionism, consciousness, and the computational mind. In: *Connectionism and the philosophy of mind*, ed. T. Horgan & J. Tienson. Kluwer. [DL]
Loar, B. (1990) Phenomenal states. In: *Philosophical perspectives, 4: Action theory and philosophy of mind*, ed. J. Tomberlin. Ridgeview. [NB]
Lockhart, R. S. (1989) The role of theory in understanding implicit memory. In: *Implicit memory: Theoretical issues*, ed. S. Lewandowsky, J. C. Dunn & K. Kirsner. Erlbaum. [EMR]
Lycan, W. (1987) *Consciousness*. Bradford Books/MIT Press. [WGL]
(1990) What is the "subjectivity" of the mental? In: *Philosophical perspectives*, vol. 4, ed. J. Tomberlin. Ridgeview. [WGL]
MacKay, D. M. (1985) Do we "control" our brains? *Behavioral and Brain Sciences* 8(4):546–47. [aDCD]
Marcel, A. J. (1983) Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual process. *Cognitive Psychology* 15:238–300. [aDCD, MV]
(1986) Consciousness and processing: Choosing and testing a null hypothesis. *Behavioral and Brain Sciences* 9(1):40–41. [aDCD]
(in press) Slippage in the unity of consciousness. In: *Perception without awareness: Cognitive, clinical and social perspectives*. Guilford Press. [rDCD]
McCloskey, D. L., Colebatch, J. G., Potter, E. K. & Burke, D. (1983) Judgements about onset of rapid voluntary movements in man. *Journal of Neurophysiology* 49:851–63. [BL]
McGinn, C. (1991) *The problem of consciousness*. Blackwell. [NB]
McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature* 264:746–48. [AWY]
Mellor, H. (1981) *Real time*. Cambridge University Press (Cambridge). [arDCD, RVG]
Merikle, P. M. (1982) Unconscious perception revisited. *Perception & Psychophysics* 31:298–301. [EMR]
Merikle, P. M. & Cheesman, J. (1986) Consciousness is a "subjective" state. *Behavioral and Brain Sciences* 9(1):42. [aDCD]
Merikle, P. M. & Reingold, E. M. (1990) Recognition and lexical decision without detection: Unconscious perception? *Journal of Experimental Psychology: Human Perception and Performance* 16:574–83. [EMR]
(1991) Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17:224–33. [EMR]
Minsky, M. (1985) *The society of minds*. Simon & Schuster. [aDCD]
Mishkin, M. & Appenzeller, T. (1987) The anatomy of memory. *Scientific American* 256:62–71. [RAMG]
Mortenson, C. (1985) Conscious decisions. *Behavioral and Brain Sciences* 8:548–49. [aDCD]
Nagel, T. (1974) What is it like to be a bat? *Philosophical Review* 83:435–45. [aDCD, NB]
Navon, D. (1991) The function of consciousness or of information? *Behavioral and Brain Sciences* 14(4):690–91. [MV]
Neisser, U. (1967) *Cognitive Psychology*. Appleton-Century-Crofts. [aDCD]
(1976) *Cognition and reality*. Freeman. [rDCD]
(1981) John Dean's memory: A case study. *Cognition* 9:1–22. [aDCD]
Neumann, O. (1990) Some aspects of phenomenal consciousness and their possible functional correlates. Presented at the conference, The phenomenal mind – how is it possible and why is it necessary? Center for Interdisciplinary Research (ZiF), Bielefeld, May 14–17. [aDCD]
Newell, A., Rosenbloom, P. S. & Laird, J. E. (1989) Symbolic architectures for cognition. In: *Foundations of cognitive science*, ed. M. Posner. MIT Press. [aDCD]
Newsome, W. T., Mikami, A. & Wurtz, R. H. (1986) Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. *Journal of Neurophysiology* 55:1340–51. [ALR]
Nicolis, J. S. (1986) *Dynamics of hierarchical systems*. Springer-Verlag. [RAMG]
Nisly, S. & Wasserman, G. S. (1989) Intensity dependence of perceived duration: Data, theories, and neural integration. *Psychological Bulletin* 106:483–96. [GSW]
Norman, D. A. (1967) Temporal confusions and limited capacity processors. *Acta Psychologica* 27:293–97. [rDCD]
Norman, D. & Shallice, T. (1980) Attention to action: Willed and automatic control of behavior. *Technical Report 8006*. Center for Human Information Processing, University of California, San Diego (La Jolla). [BB]
Ornstein, R. (1977) *The psychology of consciousness*. Harcourt Brace Jovanovitch. [MJF]
Pagels, H. (1988) *The dreams of reason*. Simon & Schuster. [aDCD]
Penfield, W. (1975) *The mystery of the mind: A critical study of consciousness and the human brain*. Princeton University Press. [MV]
Penfield, W. & Jasper, H. (1954) *Epilepsy and the functional anatomy of the human brain*. Little Brown. [aDCD]
Penrose, R. (1989) *The Emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford University Press (Oxford). [aDCD]
(1990) Précis of *The Emperor's new mind*. *Behavioral and Brain Sciences* 13(4):643–706. [GSW]
Piantanida, T. (1985) Temporal modulation sensitivity of the blue mechanism: Measurements made with extraretinal chromatic adaptation. *Vision Research* 25:1439–44. [ALR]
Piantanida, T. & Larimer, J. (1989) The impact of boundaries on color: Stabilized image studies. *Journal of Imaging Technology* 15:58–63. [ALR]
Pöppel, E. (1985) *Grenzen des Bewusstseins*. Deutsche Verlags-Anstal. (Translated as *Mindworks: Time and conscious experience*. Harcourt Brace Jovanovich, 1988.) [aDCD]
Popper, K. R. & Eccles, J. C. (1977) *The self and its brain*. Springer-Verlag. [aDCD]
Port, R. (1990) Representation and recognition of temporal patterns. *Connection Science* 2:151–76. [rDCD]
Port, R. & Anderson, S. (1990) Dynamic network models for audition. *Technical Report ISHC-TR01-RP-90*. Indiana University Institute for the Study of Human Capabilities. [rDCD]
Potter, M. (1975) Meaning in visual search. *Science* 187:965–66 [NB]
(1976) Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory* 2(5):509–22 [NB]
Pribram, K. H. (1971) Languages of the brain: Experimental paradoxes and principles in neuropsychology, 2nd ed. Prentice-Hall. [MV]
(1986) The cognitive revolution and mind/brain issues. *American Psychologist* 41:507–20. [MV]
Prinz, W. (1991) Why don't we perceive our brain states? Presented at the Seminar on Perception and Action, King's College Research Center, Cambridge, England, November, 1990. [BB]
Putnam, H. (1965) Brains and behavior. In: *Analytical philosophy, second series*, ed. R. J. Butler. Basil Blackwell. [WGL]

Pylyshyn, Z. (1979) Do mental events have durations? *Behavioral and Brain Sciences* 2(2):277–78. [aDCD]

Reingold, E. M. & Merikle, P. M. (1988) Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics* 44:563–75. [EMR]

(1990) On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language* 5:9–28. [aDCD, EMR]

Richardson-Klavehn, A. & Bjork, R. A. (1988) Measures of memory. *Annual Review of Psychology* 39:475–543. [EMR]

Ringo, J. L. (1985) Timing volition: Questions of what and when about W. *Behavioral and Brain Sciences* 8(4):550–51. [aDCD]

Roberts, M. & Summerfield, Q. (1981) Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics* 30:309–14. [AWY]

Robins, C. & Shepard, R. N. (1977) Spatio-temporal probing of apparent rotational movement. *Perception and Psychophysics* 22:12–18. [RNS]

Rosenthal, D. (1986) Two concepts of consciousness. *Philosophical Studies* 49:3. [NB]

(1990) A theory of consciousness, Report, Center for Interdisciplinary Research (ZiF), Research Group on Mind and Brain, University of Bielefeld, Germany. [DMR]

(1991) The independence of consciousness and sensory quality. In: *Proceedings of the 1989 Joint Conference of Sociedad Filosofica Ibero-Americana and Sociedad Argentina de Analisis Filosofico, Buenos Aires*, ed. E. Villanueva. Ridgeview. [DMR]

Ryle, G. (1949) *The concept of mind*. Hutchison. [aDCD]

Salter, D. (1989) Voluntary process and the readiness potential: Asking the right questions. *Behavioral and Brain Sciences* 12(1):181–82. [aDCD]

Salzman, C. D., Britten, K. H. & Newsome, W. T. (1990) Cortical microstimulation influences perceptual judgements of motion direction. *Nature* 346:174–77. [ALR]

Sarris, V. (1989) Max Wertheimer on seen motion: Theory and evidence. *Psychological Research* 51:58–68. [aDCD]

Sartre, J. P. (1943) *L'être et le néant*. Gallimard. [aDCD]

Schacter, D. (1989) On the relation between memory and consciousness: Dissociable interactions and conscious experience. In: *Varieties of memory and consciousness: Essays in honor of Endel Tulving*, ed. H. L. Roediger III & F. I. M. Craik. [NB]

Schacter, D. L., McAndrews, M. P. & Moscovitch, M. (1988) Access to consciousness: Dissociations between implicit and explicit knowledge in neuropsychological syndromes. In: *Thought without language*, ed. L. Weiskrantz. Oxford University Press (Oxford). [AWY]

Searle, J. R. (1991) Consciousness, explanatory inversion and cognitive science. *Behavioral and Brain Sciences* 13:585–641. [EMR]

Shepard, R. N. (1981) Psychophysical complementarity. In: *Perceptual organization*, ed. M. Kubovy & J. Pomerantz. Erlbaum. [RNS]

(1984) Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review* 91:417–47. [RNS]

(1987) Evolution of a mesh between principles of the mind and regularities of the world. In: *The latest on the best: Essays on evolution and optimality*, ed. J. Dupré. MIT Press/Bradford Books. [RNS]

(1989) Internal representation of universal regularities: A challenge for connectionism. In: *Neural connections, mental computation*, ed. L. Nadel, L. A. Cooper, P. Culicover & R. M. Harnish. MIT Press/Bradford Books. [RNS]

(1990) Postscript: On understanding mental images. In: *Images and understanding*, ed. H. Barlow, C. Blakemore & M. Weston-Smith. Cambridge University Press (Cambridge). [RNS]

Shepard, R. N. & Cooper, L. A. (1982) *Mental images and their transformations*. MIT Press/Bradford Books. [RNS]

Shepard, R. N. & Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science* 171:701–03. [RNS]

Sherrington, C. S. (1934) *The brain and its mechanism*. Cambridge University Press (Cambridge). [rDCD]

(1940) *Man on his nature*. Cambridge University Press (Cambridge). [BL]

Shevrin, H. & Dickman, S. (1980) The psychological unconscious: A necessary assumption for all psychological theory? *American Psychologist* 35:421–34. [EMR]

Shoemaker, S. (1981) The inverted spectrum. *The Journal of Philosophy* 74:7. [NB]

Snyder, D. M. (1988) On the time of a conscious peripheral sensation. *Journal of Theoretical Biology* 130:253–54. [aDCD]

Sperling, G. (1960) The information available in brief visual presentations. *Psychological Monographs* 74, no. 11. [aDCD]

Stamm, J. S. (1985) The uncertainty principle in psychology. *Behavioral and Brain Sciences* 8(4):553–54. [aDCD]

Stevens, S. S. (1975) *Psychophysics: Introduction to its perceptual, neural and social prospects*. John Wiley. [RT]

Stich, S. (1978) Belief and subdoxastic states. *Philosophy of Science* 45(4):499–518. [NB]

Stone, S. A. (1926) Prior entry in the auditory-tactual complication. *American Journal of Psychology* 37:284–87. [MT]

Stone, T. & Davies, M. (in press) Cognitive neuropsychology and the philosophy of mind. *British Journal for the Philosophy of Science*. [AWY]

Suga, N. (1988) Auditory neuroethology and speech processing: Complex-sound processing by combination-sensitive neurons. In: *Auditory function: Neurobiological bases of hearing*, ed. G. M. Edelman, W. E. Gall & W. M. Cowan. John Wiley [ALR]

Taylor, J. L. & McCloskey, D. I. (1990) Triggering of preprogrammed movements as reactions to masked stimuli. *Journal of Neurophysiology* 63:439–46. [rDCD, BL]

Teghtsoonian, R. (1971) On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review* 78:71–80. [RT]

(1974) On facts and theories in psychophysics: Does Ekman's law exist? In: *Sensation and measurement: Papers in honor of S. S. Stevens*, ed. H. R. Moskowitz, B. Scharf & J. C. Stevens. Reidel. [RT]

Teranishi, R. (1977) Critical rate for identification and information capacity in hearing system. *Journal of the Acoustical Society of Japan* 33:136–43. [RMW]

Ternus, J. (1926/1938) Experimentelle Untersuchung über phänomenale Identität. *Psychologische Furschung* 7:81–136. Reprinted as The problem of phenomenal identity. In: *A sourcebook of gestalt psychology*, ed. W. D. Ellis. Routledge & Kegan Paul. [MT]

Thomas, I. B., Cetti, R. P. & Chase, P. W. (1971) Effect of silent intervals on the perception of temporal order of vowels. *Journal of the Acoustical Society of America* 49:84. [RMW]

Treisman, M. (1963) Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs* 77(Whole no. 576):1–40. [MT]

(1964) The effect of one stimulus on the threshold for another: An application of signal detectability theory. *British Journal of Statistical Psychology* 17:15–35. [MT]

(1984) Temporal rhythms and cerebral rhythms. In: *Timing and time perception*, ed. J. Gibbon & L. Allan. *Annals of the New York Academy of Sciences* 423:542–65. [MT]

Treisman, M., Faulkner, A., Naish, P. & Brogan, D. (1990) The internal clock: Evidence for a temporal oscillator underlying time perception with some estimates of its characteristic frequency. *Perception* (in press). [MT]

Treisman, M. & Howarth, C. I. (1959) Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quarterly Journal of Experimental Psychology* 11:129–42. [MT]

Uttal, W. R. (1973) *The psychobiology of sensory coding*. Harper & Row. [GSW]

(1978) *The psychobiology of mind*. Erlbaum. [MV]

(1979) Do central nonlinearities exist? *Behavioral and Brain Sciences* 2(2):286. [aDCD]

Van der Waals, H. G. & Roelofs, C. O. (1930) Optische Scheinbewegung. *Zeitschrift für Psychologie und Physiologie des Sinnesorgane* 114:241–88. (Also [1931] 115:91–190.) [arDCD]

Van Gulick, R. (1985) Conscious wants and self-awareness. *Behavioral and Brain Sciences* 8:555–56. [aDCD]

(1991) Consciousness may still have a processing role to play. *Behavioral and Brain Sciences* 14(4):699–700. [MV]

(forthcoming) Understanding the phenomenal mind: Are we all just armadillos? In: M. Davies & G. Humphreys. *Consciousness: A mind and language reader*. Blackwell. [NB]

Velmans, M. (1991) Is human information processing conscious? *Behavioral and Brain Sciences* 14(4):651–669. [aDCD]

(1991a) Consciousness from a first-person perspective. *Behavioral and Brain Sciences* 14(4):702–19. [MV]

Vendler, Z. (1972) *Res cogitans*. Cornell University Press. [aDCD]

(1984) *The matter of minds*. Clarendon Press. [aDCD]

Volpe, B. T., LeDoux, J. E. & Gazzaniga, M. S. (1979) Information processing of visual stimuli in an "extinguished" field. *Nature* 282:722–24. [MJF]

von der Malsburg, C. (1987) Synaptic plasticity as basis of brain organization. In: *The neural and molecular bases of learning*, ed. J.-P. Changeux, M. Konishi, D. Konferenzen & S. Bernhard. John Wiley. [rDCD, ARD]

Wallace, M. A. & Farah, M. J. (in press) Savings in relearning as a measure of "covert recognition" in prosopagnosia. *Journal of Cognitive Neuroscience*. [MJF]

Warren, R. M. (1974a) Auditory temporal discrimination by trained listeners. *Cognitive Psychology* 6:237–56. [RMW]
(1974b) Auditory pattern discrimination by untrained listeners. *Perception & Psychophysics* 15:495–500. [RMW]
(1981) Measurement of sensory intensity. *Behavioral and Brain Sciences* 4:175–223. [RMW]
Warren, R. M. & Ackroff, J. M. (1976) Two types of auditory sequence perception. *Perception & Psychophysics* 20:387–94. [RMW]
Warren, R. M., Bashford, J. A., Jr. & Gardner, D. A. (1990) Tweaking the lexicon: Organization of vowel sequences into words. *Perception & Psychophysics* 47:423–32. [RMW]
Warren, R. M., Gardner, D. A., Brubaker, B. S. & Bashford, J. A., Jr. (1991) Melodic and nonmelodic sequences of tones: Effects of duration on perception. *Music Perception* 8:277–90. [RMW]
Warren, R. M. & Obusek, C. J. (1972) Identification of temporal order within auditory sequences. *Perception & Psychophysics* 12:86–90. [RMW]
Warren, R. M. & Warren, R. P. (1968) *Helmholtz on perception: Its physiology and development.* John Wiley. [RMW]
Wasserman, G. S. (1985) Neural/mental chronometry and chronotheology. *Behavioral and Brain Sciences* 8(4):556–60. [aDCD, GSW]
Wasserman, G. S. & Kong, K.-L. (1979) Absolute timing of mental activities. *Behavioral and Brain Sciences* 2(2):243–304. [aDCD, GSW]
Wasserman, G. S., Wang-Bennett, W. T. & Miyamoto, R. T. (1990) Temporal dispersion in natural receptors and pattern discrimination mediated by artificial receptors (cochlear implants). In: *Psychophysical explorations of mental structure*, ed. H.-G. Geissler, M. H. Müller & W. Prinz. Hogrefe & Huber. [GSW]
Weiskrantz, L. (1986) *Blindsight: A case study and implications.* Oxford University Press (Oxford). [AWY]
Welch, R. B. (1978) *Perceptual modification: Adapting to altered sensory environments.* Academic Press. [aDCD]
Wertheimer, M. (1912) Experimentelle studien über das Sehen von Bewegung. *Zeitschrift für Psychologie* 61:161–265. [aDCD]
Young, A. W. & de Haan, E. H. F. (1990) Impairments of visual awareness. *Mind and Language* 5:29–48. [AWY]
Zak, M. (1991) Terminal chaos for information processing in neurodynamics. *Biological Cybernetics* 64:343–351. [RAMG]

Please send BBS your email address

All BBS Associates and any non-Associates who have served as referees, commentators or authors, or who are qualified and interested in serving as referees or commentators for BBS:

Please send us your electronic mail address, if you have one. (If you don't have one, you are strongly urged to look into the advantages of getting one -- not only for BBS's sake!)

BBS is implementing more and more of its peer communication functions by electronic mail. This not only increases the speed and efficiency of BBS's interaction with the world biobehavioral and cognitive science community, but it dramatically increases its scope and range as well. Abstracts can be circulated by email in advance to allow potential commentators to nominate themselves. Referee reports can be submitted by email. The BBS Associateship can be more representatively canvassed to determine what topics and authors they would like to see treated in BBS. New Associates can be nominated by email, etc.

Electronic mail addresses can be sent to our regular mail address. (Any available *departmental* or *institutional email directories* would be very helpful too.)

Behavioral and Brain Sciences
20 Nassau Street, Room 240
Princeton NJ 08542

Or they can be sent by electronic mail to any of the following electronic mail addresses. (Because email is not yet reliable, please try several until you receive confirmation that your message has been received.)

INTERNET

harnad@confidence.princeton.edu
harnad@princeton.edu
harnad@elbereth.rutgers.edu
srh@flash.bellcore.com

BITNET

harnad@pucc.bitnet

UUCP

harnad@princeton.uucp
princeton!confidence!harnad

CSNET

harnad%confidence.princeton.edu@relay.cs.net

Along with your email address you are encouraged to include you suggestions about current BBS editorial policy and directions you would like to see BBS take in the future.