The Journal of Mind and Behavior
Autumn 1999, Volume 20, Number 4
Pages 421–436
ISSN 0271–0137

A Testable Mind–Brain Theory

Ralph L. Smith

Tucson, Arizona

Proceeding from the observation by Ryle (1949/1984) that I cannot prepare myself for the next thought that I am going to think, I argue that conscious acts cannot control my bodily motions or thoughts. This position is not compatible with indeterminism. I also argue that consciousness represents the irreducible and multi-modal output (across a hypothetical brain–consciousness interface) of the behavioral control system *sensors* necessary for the control of human behavior demonstrated by Marken (1988). My analysis supports one experimental result obtained by Libet, Gleason, Wright, and Pearl (1983), namely, that the initiation of a "voluntary act" is an unconscious cerebral process. I conclude that the following are not realizable: "mental intentions" acting on the supplementary motor area as postulated by Eccles and Robinson (1984), and "veto," a conscious abort of a motor act after subjects reported "wanting to act" (Libet, 1985). These two items would seem to be amenable to test by studies similar to or refinements of Libet's.

This paper is an examination of the relation between mind and brain. That such relation is one of exceptional complexity is evidenced by the extraordinary diversity of disciplines involved; the duration of time (hundreds, even thousands, of years) over which the problem has been debated; the profusion of so-called mind–brain (or mind–body) theories;[1] and the

The author wishes to thank Richard S. Marken for helpful discussion on the subject of controlled events in human behavior. Requests for reprints should be sent to Ralph L. Smith, Ph.D., 2740 North Tomahawk Trail, Tucson, Arizona 85749.

[1]We never speak of a "consciousness–brain" theory, yet consciousness is often the center of attention in mind–brain theories. Velmans (1996, p. 2) states that, "In some writings, 'consciousness' is synonymous with 'mind.' However, given the extensive evidence for nonconscious mental processing . . . this definition of consciousness is too broad." Searle (1984, p. 10) states, "By 'mind' I just mean the sequences of thoughts, feelings, and experiences, whether conscious or unconscious, that go to make up our mental life. But the use of the noun 'mind' is dangerously inhabited by the ghosts of old philosophical theories." Later, Searle (pp. 15–17) identifies four features of the mind: consciousness, intentionality, subjectivity, and mental causation. However one defines the mind–brain problem, I shall be concerned herein specifically with consciousness.

limited success attained therefrom until the twentieth century, largely because of the unavailability of sensitive experimental techniques before then. To no small degree, progress is hampered by differences in cross-disciplinary terminology.

Although consciousness cannot be examined by itself in the laboratory, there is general, but not universal, agreement that it is related to what we physically do. Virtually nothing is known about its ontology. Much has been learned epistemologically, certainly enough that the study of consciousness has finally been accepted as a respectable occupation. Supplementing, or guided by, laboratory investigations, theoretical papers of increasing sophistication issue from many academic fields on various aspects of the mind–brain problem. Such aspects include, in part, classic subjects such as whether or not we have free will (i.e., Is human behavior deterministic or indeterministic?); whether or not consciousness belongs to the physical world; and whether or not it could interact with our brains should it not so belong. Bunge (1977/1980) provided a formal summary and description of several candidate mind–brain relationships, discussing the merits of each but noting that there are underlying obstacles to drawing any universally acceptable conclusions:

> This commentary deals with the so-called mind–body problem. This is the set of questions about the nature of the mental and its relations to the bodily . . . . These questions are rather difficult to answer. However, I submit that the difficulty is not wholly intrinsic but has been compounded by hurdles such as the following. Firstly, several doctrines concerning the mind–body problem have some ideological basis or other — and ideologies are not particularly interested in fostering conceptual clarity and empirical investigation. Secondly, the very formulation of the mind–body problem employs certain concepts, such as those of substance, emergent property, state and event, which are far from clear. (p. 633)

To the last sentence above, I would add the concept of agency, a hypothetical power assuring that, unless physically constrained, I "could have done otherwise."

Arguments are presented herein that, collectively, constitute a mind–brain theory to the extent that theories can be tested against existing experimental data and can predict the outcome of future experiments. Given reasonable assumptions, the development leads to the conclusions that (a) consciousness performs an important function in the control of human behavior, and (b) consciousness cannot act in the philosophical sense. Judging from the literature (e.g., Taylor, 1966/1973), the concept of "doing" or "acting" in the philosophical sense is far from elementary.[2] The notion that you or I can

[2]Where the word forms "do" and "act" appear in individual quotes ("do," "act," 'do,' 'act'), the modifier "in the philosophical sense" is to be understood.

"do" or "act" is not taken for granted herein and care must be exercised in the interpretation of familiar language such as "I thought such and such" or "I did this or that."

Clearly, the doctrines of determinism and indeterminism cannot both be correct. Any information, however small, that might favor one over the other should not be overlooked. As a case in point, the unexpected experimental results of Libet, Gleason, Wright, and Pearl (1983) fortuitously enable qualitative testing of the mind–brain theory proposed herein. As Dennett (1991) writes:

> [Libet] 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 in the 300–500msec range. This is huge — up to half a second — and it does look ominous to anyone committed to the principle that our conscious acts *control* our bodily motions. (p. 164)

The *feeling* that we can behave (do, think) as we wish and *could* have chosen differently, "consciously controlling" our nonautonomic and nonreflex bodily motions and our thoughts, is almost unshakeable. Even to consider otherwise is incomprehensible to the average person and heresy to many indeterminists. This all-pervasive feeling (the subjective sense of free will) has penetrated our very manner of speaking: "I chose this or that," and so on. If, in addition to this *feeling*, it should evolve that we *can* consciously control our behavior and *could* have chosen differently, then it could be said that we have true free will, not just the subjective sense of free will.

The concept of an "I" that can "consciously control" is not clear if controlling means "what a control *system* does." *Whatever* is being controlled (e.g., a guided weapon, a biological process such as glandular secretion, or human behavior), several well-defined logical elements are involved.[3] In the control of human behavior, consciousness is at best only *one* of those elements. I conclude that, whatever its ontology, consciousness alone cannot control *anything*. Stronger arguments supporting this conclusion are offered below.

Because this paper is directed at a study of consciousness, which historically is especially resistant to definition, an attempt is now made to attend to terminology. As Güzeldere (1995) writes:

> The first difficulty in the study of consciousness is the conceptual tangle encompassing the subject matter. There are many meanings of the term "consciousness," and many variations within each distinct concept . . . . The many different senses of consciousness have to be teased apart; its many faces individually revealed. (p. 33)

---

[3]The logical elements of control are discussed by many authors (e.g., DiStefano, Stubberud, and Williams, 1990).

Sommerhoff (1996, p. 142) provides what he represents as a "dictionary average," defining consciousness as "an awareness of the surrounding world, of the self, and of one's thoughts and feelings." This, of course, assumes we already agree on what "self" means. That assumption is unfortunately unjustified. Conflicting interpretations of a (cognitive as opposed to physical) "self" are easy to find. Strawson (1997, p. 406) does not ". . . want to exclude in advance the view that there is no such thing as the self . . . ." He subsequently refers to a "sense of self," "sense" suggesting that at least some part of self is conscious experience. Strawson (p. 408) proposes that mental self is "*ontically distinct* from all other things" (further admitting that "thing" is also in need of defining). However, he adds (p. 409) that such a proposal needs careful qualification if it is to survive. Perlis (1997, p. 509) *identifies* consciousness with self, thus making the use of "self" superfluous. James (1892/1984) clearly attributed dual characteristics to self: the empirical ego (the "known," see p. 159) and the pure ego (the "knower," see p. 175). The pure ego, the "I," is conscious; the empirical ego is the "me," *one* of the things of which the pure ego is conscious. Eccles (1953) provided a simple definition that captures the essence of self:

> ["Self" connotes] a unity that derives from a linking by memory of conscious states that are experienced at widely different times — spread over a life-time. Thus, in order that a "self" may exist, there must be some continuity of mental experiences and, particularly, continuity bridging gaps of unconsciousness. (p. 264)[4]

Although Eccles's description most accurately agrees with my idea of self, "self" is a metaphysical (beyond the physical) concept, and so is subject to personal interpretation. Furthermore, my examples are by no means all-inclusive. Therefore, for the purposes at hand, and considering the problems in defining self as it relates to consciousness, I find it sufficient to take my lead from Crick (1995, p. 20), who writes, "Everyone has a rough idea of what is meant by consciousness. It is better to avoid a *precise* definition of consciousness because of the dangers of a premature definition." Each person's instinctive concept of consciousness should be sufficient for understanding this paper but, not wishing to deny its subtleties, I refer the interested reader to Natsoulas (1996, pp. 284–285) for a bibliography of his analyses of several definitions of consciousness in the Oxford English Dictionary.

---

[4]The gaps referred to by Eccles are not confined to sleep. As James (1890/1950, p. 237) points out in discussing the continuity of thought, consciousness is observed to have random "time-gaps" during which it disappears altogether. Strawson (1997, p. 422) agrees: ". . . my fundamental experience of consciousness is one of *repeated returns into consciousness from a state of complete, if momentary, unconsciousness*."

The word "I" is fraught with similar uncertainty partly because of its unclear association with self. Deikman (1996) concentrates on a cognitive "I" he identifies with awareness in a way subtly different from his concept of self. Strawson (1997, p. 408) acknowledges the potential problem in distinguishing between the "public I" and the "private I" associated with a mental self.

To minimize confusion when using the unsubscripted and unmodified personal pronoun "I" henceforth, I introduce I_0 simply as an abbreviation for "my consciousness" and reserve the word "I" strictly for use in the public, conversational sense. That particular "I" is specifically *not* to be interpreted as either I_0 or "private I." It refers to the undifferentiated totality of my body and mind. The precise meaning of "private I" is made clear below.

The main purpose of this paper is the analysis of I_0 and its relation to my behavior. What is the connection, if any, between consciousness and behavior? In one of his later books, Penfield (1975, p. 21) recalls the first occasion (1933) on which one of his epileptic patients reported a "flashback" in response to electrical stimulation of the cerebral cortex: he "re-lived" (Penfield's word) all that he had been aware of at an earlier time. As Penfield wrote of a subsequent patient:

> D.F. could hear instruments playing a melody. I re-stimulated the same point thirty times (!) trying to mislead her, and dictated each response to a stenographer. Each time I re-stimulated, she heard the melody again. It began at the same place and went on from chorus to verse. When she hummed an accompaniment to the music, the tempo was what would have been expected. (p. 22)

The phrase "cerebral (or neural) correlates of conscious experience" has come into general use to denote the relation between the electrical stimulation of the brain and the conscious responses that result. Such correlates constitute compelling evidence that there is an ontological relation between the brain and consciousness, yet there is no indication that consciousness belongs to the physical world. I shall accordingly follow Searle (1994, p. 116), who argues that consciousness is irreducible, meaning that I_0 is *not* part of the physical world in general and my brain in particular. Nevertheless, there appears to be *some* ontic connection between brain and consciousness which, for lack of a better term, I shall refer to as the "brain–consciousness interface." It is not obvious that such an interface would require modifying the laws of physics everywhere: some latent property of the physical world may exist enabling communication between brain and consciousness at their interface but having no effect elsewhere.

In addition to the difficulties in defining consciousness, debates still continue on what function consciousness performs, if any. As might be suspected, many positions can be taken on this issue, leading to a number of

cognitive theories ("isms"), each with its own set of modifiers (e.g., emergentist materialism). Of particular interest in the study of the function of consciousness is a paper by Elitzur (1989), who concludes that consciousness is causal. I concur as long as Elitzur's "causal" means that consciousness plays a rôle in our behavior, *provided* that such rôle does not include any power of consciousness to *control*. In general, however, I tend to avoid the terms "cause" and "effect," following Margenau (1950, chap. 19).

A Logical Premise and Some Implications Thereof

In this section I propose a mind–brain theory evolving from a logical premise that should be easy both to verify if correct and to refute if wrong. Any extent to which an appeal to logic is possible reduces the need to depend on intuition. The analysis leads to a critical question having at least three possible answers, the most likely one (in the author's opinion) strongly suggesting an origin of consciousness and leading to a rational hypothesis as to its function.

To establish the premise, I begin by examining a logical truth apparently first noted by Ryle (1949/1984). I reach the conclusion that I_0 cannot "act," meaning that I_0 has no power to *originate* any physical *or mental* activity. Being unable to "act" does not, however, imply "serving no purpose": an essential rôle for I_0 in the control of human behavior will be discussed in the next section.

First, suppose that, as I write these words, I am trying to think of the next word to write. With Baars (1997, p. 49), I am occasionally conscious of a "waiting period" for the "next word that will appear, but has not yet appeared, in I_0" (sometimes referred to below simply as "next word" where no confusion is likely to arise). During this period, an *unconscious* "incubation" process takes place ending when that next word appears in I_0, at which time it loses its status as "next word." In common parlance, and recalling my use of the unsubscripted "I," "I thought of that word." That example of common parlance, however, must *not* be misread as equivalent to "I_0 thought of that word." It is essential to realize that those two statements have quite different connotations. For well-memorized material or in normal conversation, the waiting period may be so short as to be unnoticeable.

Second, instead of using the observational approach above, I might ask the *logical* question: Is it possible for me to know (be conscious of) that next word? The answer is trivially "No" because of the way "next word" is defined, namely, a word that has *not yet appeared* in I_0. One might incorrectly argue that if I were, say, reciting some familiar verse, surely I would know the next word at any point in the verse even though it had not actually appeared in I_0. In truth, although the next word is certainly in my long-term memory, it is,

by definition of "next," logically impossible for me to be conscious of it. Although his argument is somewhat different, Ryle (1949/1984, p. 197) reached essentially the same conclusion when he wrote, "One thing I cannot prepare myself for is the next thought that I am going to think."

Thought, however, is not always easily divided into chunks in such a way that "next thought" has a precise meaning: a thought may entail modalities such as sight and sound that have continuous qualities. Words are more easily accepted than protracted thoughts as discrete, countable entities, but even words can be broken down into syllables and the syllables subdivided into fragments of a sound wave. What, for example, might I mean by my "next feeling" or "next intention" if such feelings and intentions are continuous? To resolve the question, it is sufficient to generalize the meaning of "next" so as to refer to intervals of time, dt, as small as we like when dealing with conscious experiences that are inherently continuous. Such a position recognizes that the original problem lay not in the intrinsic meaning of "feeling" or "intention" but in the fact that each is continuous in time. I therefore state as my logical premise ("X" for short) that:

X: It is impossible for me to know my next conscious experience.[5]

As explained above, "next" always carries the connotation "not yet in I_0."

In retrospect, X might appear obvious and inconsequential. However, it raises a critical question not asked by Ryle: If I cannot know my next conscious experience, *what determines what it will be?* Short of (1) attributing self-generating powers to I_0 (which would make my body superfluous) or (2) introducing metaphysical concepts (which are highly controversial), the most plausible alternative I see is that (3) my next conscious experience is generated by *processes in my brain*, of which I am not conscious because, given Searle's irreducibility argument, I_0 is *not* part of my brain. The generation of that next conscious experience must take place at the brain–consciousness interface.

Alternative (3) is equivalent to the statement that I_0 at *any* time is determined by processes in my brain that took place *earlier* but of which I was not conscious. Either way, the argument supports Crick's (1995) "second assumption":

> My second assumption was that one is not conscious of the "computations" done by this part of the brain but only of the "decisions" it makes — that is, its plans. (p. 266)

[5]It *is* possible for me to be conscious of an *expectation* as to what my next conscious experience will be.

While I may *feel* that I_0 somehow creates words as if it were actively "thinking," I claim that I_0 has no such capability. Being directly generated by my brain, as I contend, I_0 itself cannot consciously "do" anything such as "steer" my thoughts, themselves part of I_0, along certain lines so that thoughts consistent with some "desired subject" or "desired goal" appear in I_0. If I subjectively feel, as we all do, that I_0 controls my thoughts and behavior, it can only be because that feeling is itself a part of I_0 generated by my brain.

Although one might think of I_0 as information in the computer sense, such an association would be misleading because I_0 is not something that is processable by a computer, whether analog or digital. Calling it a "data file" in the computer sense is also misleading because, as stated above, there is no indication that I_0 belongs to the physical world. With no knowledge as to the ontology of consciousness, it is futile to discuss how many dimensions might be associated with I_0, where it might be "stored," or whether it has any extension whatsoever. To indicate metaphorically what happens at the brain–consciousness interface, I look upon I_0 as a "virtual file" that is "written" by my brain. Here, "virtual" is used as opposed to "real" (in the physical sense) to emphasize the position that I_0 is irreducible.

Baars (1997, p. 42) and Chalmers (1996, pp. 6–11) provide descriptions of the conscious world. It is neither appropriate nor necessary to examine every possible ingredient of consciousness: the following limited discussion should be sufficient to test the credibility of X.

Consider the conscious experience associated with motor activity. As everyone should be able to verify, we are not conscious of the *efferent* (outgoing) nerve signals that activate our muscles. When I move my arm, my consciousness consists of at least the following two components: (1) the *intention* (generated by my brain) to move my arm, and (2) the consciousness resulting from the conversion, at the brain–consciousness interface, of *afferent* (incoming) signals from my eyes, skin, tendons, muscles and joints as my arm undergoes flexion or extension. It should be apparent that conscious experience resulting from my brain's interpretation of virtually all *afferent* signals (exteroceptive, or "outside world"; interoceptive, or viscera and glands; and proprioceptive, or muscles and tendons) cannot be anything that I_0 causes or brings about: I_0 is a *consequence* of those afferents. However, X is not limited to such "perceptions": it applies to *all* conscious experience, so that "voluntary (i.e., conscious) control" of cognitive processes is also something of which I_0 is not capable. Baars's (1997) remarks support this conclusion:

> If we keep in mind the fact that you and I have no idea how we control our muscles, a complex and subtle process that is not available to consciousness, the question arises, Who or what is doing the learning in biofeedback? Common sense says that "we" are learning, as if there is some centralized self that is in control of all the details. But that cannot be true, because the everyday "we" has no access to the necessary information.

> If we think of the brain as a massive decentralized society of biocomputers, it may be more realistic to say that the motor system is learning something, *based on conscious information that is made available to many unconscious local control systems*. We are not doing the learning; they are. (p. 59)

Conscious experiences such as desires and intentions are subtler than those resulting from afferent pathways. Thus, James (1890/1950, p. 254) claims that ". . . we think afterwards we felt the very *tendencies* of the nascent images to arise, before they were actually there." I do not see any conflict between James's claim and mine. It seems reasonable that, in the process of "writing to my consciousness," the brain could generate, at the brain–consciousness interface, extremely vague and weak, but conscious, messages in the early stages of some processing, and that those "wisps of consciousness" might indeed hint at some subject matter or train of thought that, although not fully formed, is imminent. Nevertheless, those wisps were conscious experiences, just as are the future conscious experiences they portend.

In summary, I contend that I_0 itself cannot be "observed": there *is* no "observer," no "ghost in the machine." I conclude that I consist entirely of my body and I_0. The "private I" mentioned above *becomes* I_0. There *is* no addition to I_0 that "feels feelings" or "thinks thoughts" or "is conscious of being conscious." Any attempt to enlarge upon the meaning of my "private I" can serve no purpose because it would have to be built on concepts that are not conscious.

At this point, two conclusions can be drawn based on the properties of I_0 argued above. First, an answer to a question rarely asked: Why do we have the subjective sense of free will? What precisely gives me the *feeling* that I am "free" in the sense of being able to want anything or to do anything I want unless physically constrained? The answer, I maintain, is that I cannot sense my brain telling me what to do. I_0 is *generated* by my brain but is not privy to its functioning: as Cotterill (1989, p. 271) quotes Spinoza, "Men think themselves free because they are conscious of their volitions and desires, but are ignorant of the causes by which they are led to wish and desire."

The second conclusion is somewhat subtler, requiring a brief discussion of the concept of choice. For choice to have meaning, one must have at least two alternatives. For me to make a decision, I must have some *criterion* by which to judge or else leave the selection up to chance such as by a flip of a coin, which is not "choosing." The question then is: What determines the criterion? The answer is: another choice from a *set* of criteria. The accepted term for this never-ending process is "infinite regress," considered "clearly untenable" by Daveney (1964, p. 515). However, the dilemma disappears given the properties of I_0 because I_0 cannot make choices. My brain obviously resolves the problem ultimately, although we do not yet know how. A third conclusion, independent of I_0, is that my theory is consistent with

causality: at no point have I introduced the requirement that "I could have done otherwise," the sine qua non of free will and the hallmark of the philosophical concept of agency.

A Rôle for Consciousness

The previous section evolved from the premise X and was mainly concerned with the *characteristics* of I_0. In the present section, I discuss a rôle that I_0 might play in the *control of my behavior*. Whereas the arguments in the previous section were built around an impersonal premise, some arguments in this section involve introspection and therefore are subject to personal error. I remind the reader that my brain generates the conscious experience of introspecting, not I_0 (which cannot "do" anything).

That human behavior is *controlled* is persuasively argued by Marken (1988). Here, "behavior" refers not to the autonomic control systems in the body but to what I (not I_0!) physically do. Marken (p. 197) identifies a controlled event as ". . . a physical variable (or a function of several variables) that remains stable [i.e., within defined limits] in the face of factors that should produce variability." As an example, suppose a subject uses an optical pointer to try to follow a randomly moving spot of light (the "target") on a screen. The random motion of the target is a factor that produces variability of the distance between the pointer and target spots. *Control* is demonstrated to the extent that the variance of the distance between the two spots is smaller than the variance of the distance between the target and any fixed point on the screen (corresponding to zero pointer movement).

Although the body has autonomic control systems regulating heart rate, blood pressure, temperature, glandular secretions, pupillary size, shivering and so on, that activity is not conscious experience. There is a popular biofeedback theory that we can "voluntarily control" certain autonomic functions such as blood pressure. According to the thesis of this paper, such voluntary (i.e., conscious) control is illusory because I_0 cannot "do" anything.

The situation is markedly different in the case of my behavior: I strongly *feel* that I_0 controls my physical behavior, thoughts and emotions although the thesis of this paper is that I_0 cannot exercise this control. Without violating that thesis, I can ask what rôle I_0 might play in the specialized but as yet not understood systems that must exist to control my behavior. The answer may entail a departure from the traditional applications of control theory (i.e., engineering applications, and biological systems that do *not* involve consciousness), but my approach is most emphatically not intended to be reductionist in the ontological sense (Searle, 1994, p. 113).

Arguably, the most important feature of I_0 is that, among other things, it *defines* the physical universe for *me* upon conversion, at the brain–consciousness

interface, of my exteroceptive, interoceptive and proprioceptive afferents, a barrage of neural events all appearing to be physiologically identical. In some unknown way, my brain, via the brain–consciousness interface, transforms those afferent signals into a relatively small number of *qualia* (sight, sound, touch, taste, smell, balance, and various subdivisions). Put another way, my brain *senses* macroscopic *properties* (i.e., qualia) associated with the physical world in a way similar to, but more complex than, the way the metal coil of a thermostat *senses* a macroscopic *property* (temperature) associated with the kinetic energy of the gas molecules surrounding it. Qualia, which are part of I_0, represent one type of output of my brain at the brain–consciousness interface just as the expansion or contraction of the coil represents the output of the thermostat. As an example, my visual "sensor" would include my eyes, all the neural data they generate, and all subsequent processing up to the *perception of sight*, a component of I_0 representing the *output* of my visual sensor.

I refer to any biological control systems whose sensor outputs are conscious experience as *behavioral* control systems. The need to sense (generate consciousness of) only a relatively small number of macroscopic properties makes the brain's behavioral control systems more efficient by many orders of magnitude compared to what the case would be if those systems had to process undifferentiated afferent neural data in the absence of consciousness.

Just as quanta refer to *amounts* of something, qualia refer to *qualities* of something such as its color or odor. Qualia are the basic ingredients of conscious *patterns* used by my behavioral control systems to identify objects in my environment, sources of food, warnings of danger, and myriad other components of the outside world.[6] Since I can recall innumerable spatial or temporal patterns, they must have been written into memory as some form of recoverable "cognitive information."[7] Several possibilities exist as to how this might come about. One is that my brain saves the *afferent information* and writes it to memory. This would be extremely wasteful of storage compared to that required for patterns. A second and more likely possibility is that some part of my brain, but not the same part that writes to I_0, reads the *patterns* in I_0, converts them to cognitive information by suitable processing and stores them in memory. A third, but unlikely, possibility is that I_0 *writes* the patterns to my brain in a format appropriate for cognitive storage. Is it

[6]Qualia do not always lead to pattern *recognition*. Introspection suggests that pattern recognition develops from repeated exposure to qualia associated with some specific "real-time" spatial or temporal stimulus (e.g., a visual image or a spoken word) until a durable, recallable, memory imprint is established. Recognition is the matching of the stimulus with its memory imprint.

[7]By "cognitive information," I mean data (generated by my brain, read from I_0 by my brain, and/or stored by my brain) that my brain *could* write (transform) to consciousness. I_0 is not to be interpreted as cognitive information.

reasonable to suppose that I_0 can write to my brain? Not if the metaphor of consciousness as a file, albeit virtual, is valid — files can neither read nor write. For I_0 to be able to write (metaphorically) to my brain, it would have to have computer-like properties itself, of which there is neither need nor evidence. Of the three possibilities, the second is the most plausible. I therefore envision I_0 as a virtual input/output file that my brain can both write to *and read from* across the brain–consciousness interface. The *contents* of that file constitute my consciousness, the irreducible and multi-modal output of my behavioral control sytem *sensors*.

What is added to memory, when I_0 is "read," is cognitive information. Searle (1994) explains the difference between the processing of cognitive and computer information:

> What I just imagined an opponent saying embodies one of the worst mistakes in cognitive science. The mistake is to suppose that in the sense in which computers are used to process information, brains also process information . . . . A standard computational model of vision will take in information about the visual array on my retina and eventually print out the sentence, "There is a car coming toward me." But that is not what happens in the actual biology. In the biology a concrete and specific series of electrochemical reactions are set up by the assault of the photons on the photo receptor cells of my retina, and this entire process eventually results in a concrete visual experience . . . . The "information" in the brain is always specific to some modality or other. It is specific to thought, or vision, or hearing, or touch, for example. (pp. 223–224)

If I assume, as I shall, that my brain is a collection of biological control systems, then I_0, converted to cognitive information, is *precisely the kind of information my behavioral control systems need* if I am to navigate and survive in the physical world. There are, to be sure, contents of I_0 other than a representation of my physical universe. Examples would be the elements of intentionality described by Searle (1984, p. 16).[8] I perceive qualia and the elements of intentionality as examples of *modes* of consciousness (thus my use of the term *multi-modal*), though such examples are not necessarily exhaustive.

I now summarize what I believe to be true about I_0 at this point:

1. At time t, I cannot know the next conscious experience that will appear in I_0 at time t + dt.
2. Except for James's "time-gaps" described above, and dreamless sleep, I_0 is continuous and therefore has meaning over the arbitrarily short periods of time, dt, discussed in developing X.
3. I_0 is generated by unconscious processes in my brain.

[8]The elements of intentionality described by Searle are not qualia, yet they are conscious experiences and therefore components of I_0.

4. I_0, when "read," represents the cognitive information my behavioral control systems require.
5. I_0 has no computational powers. It simply represents the "stream of consciousness" generated by my brain. However, *control* requires a control *system*, and as such requires computational powers, or their biological equivalents. I therefore conclude that I_0 *itself* cannot control anything — not itself and not any processes in my brain — although it leads to cognitive information essential to the biological systems that *do* control my behavior. I envision my entire brain as a complex of biological, often adaptive, control systems evolving in Darwinian fashion.[9]

Compatibility with Libet's Results

Libet et al. (1983) and Libet (1985) describe experiments leading to the conclusion that the initiation of a so-called "voluntary act" is actually an unconscious cerebral process. Not only does my theory confirm this conclusion qualitatively (see below), it also predicts the answers to two questions that might eventually be resolved experimentally using techniques similar to or refinements of Libet's. The first question addresses the postulate by Eccles and Robinson (1984, pp. 156, 161) of a "mental intention" acting on the supplementary motor area (SMA) to initiate observable "readiness potentials" (RPs). Such mental intention would be counter to my conclusions because it would require that I_0 (here, a mental intention) have the power to "act." The second question addresses the possibility of a conscious "veto," a hypothetical ability of Libet's subjects to *consciously abort* a previously reported "wanting to act." My theory denies that any such abort could be conscious: as mentioned above, I_0 cannot "act." The matter of veto is currently undecidable because, as Libet (1985, p. 538) writes, ". . . there is presently available no technique for recording and analyzing any RPs that may be associated with such spontaneous, irregularly appearing conscious urges to act that do not lead to an actual motor event."

What follows is a description of the technique of Libet et al. (1983), as described in the abstract:

> The recordable cerebral activity (readiness-potential, RP) that precedes a freely voluntary, fully endogenous motor act was directly compared with the reportable time (W) for appearance of the subjective experience of "wanting" or intending to act. The onset of cerebral activity clearly preceded by at least several hundred milliseconds the reported time of conscious intention to act. (p. 623)

[9]The philosophy of teleology is similar in principle to control, the "teleological goal" corresponding to the commanded state of the system controlled.

The experimental procedure consisted, in part, of (a) attaching passive scalp electrodes for the purpose of sensing cortical activity (RP), and (b) attaching electromyogram (EMG) electrodes over the muscle tissue responsible for flexion of the fingers and/or wrist for the purpose of determining when such flexion began. Three pieces of information were used: the reported time at which the subject first "wanted" to flex fingers and/or wrist; the time cortical activity was first sensed by the scalp electrodes; and the time muscle activity was first recorded on the EMG. The subjects were not directed to flex at any particular time but just whenever they "felt like it." The result of interest is that the RP appeared *before* the reported time of wanting to act. Based on the data in Table 2 (Libet et al., 1983, p. 631), "wanting to act" occurred on the average about 350 milliseconds (and occasionally more than 500 milliseconds based on the raw data of Table 1, p. 630) *after* detection of type II RPs ("no preplanning").

The above experimental result is in accord with what one would expect in view of X, which led to the conclusion that "wanting to act" (a conscious experience) should be *preceded by* related cerebral activity, although X cannot provide *quantitative* information as to the size of the delay. However, if we "consciously controlled" our muscular activity, one would expect to find that our "wanting to act" would *precede* any related cerebral activity. As Libet (1985, p. 529) writes, "If a conscious intention or decision to act actually initiates a voluntary event, then the subjective experience of this intention should precede or at least coincide with the onset of the specific cerebral processes that mediate the act."

I conclude that the *experimental* work of Libet and his coworkers is consistent with my *logical* inference, starting with X, that conscious experience is generated by processes in my brain of which I am not aware. If my arguments are correct, then (1) Libet's results are qualitatively confirmed, (2) there can be no "mental intention" acting on the supplementary motor area, and (3) no "conscious abort" is possible. Libet's experiments are capable of being repeated, refined and enlarged in scope, and items (2) and (3) either are or may become susceptible to study, enabling those predictions of the theory to be tested.

Summary and Discussion

A mind–brain theory has been presented with the following characteristics: (a) it proceeds from a logical premise, X; (b) it suggests a rôle for consciousness; (c) it implies that consciousness has no power to control; (d) it is consistent with available experimental data; (e) it is consistent with the subjective sense of free will; (f) it is consistent with causality;[10] (g) it does not

[10]Consciousness cannot change the quantum-mechanical probability wave. For supporting arguments, see Searle (1984, p. 87) and Schrödinger (1944/1993, p. 86).

have the problem of infinite regress; (h) it is not driven by ideological considerations; and (i) it does not rely on unclear concepts such as agency, substance, emergent property, state and event.

The theory I have described represents a determinism enhanced by my analysis of the function of consciousness. I believe this approach agrees with the exact sciences as much as can be, given that we have no knowledge of the ontology of consciousness. My conclusion as to free will should be evident but is by no means original: as Planck (1937, p. 105) wrote, "Looked at objectively, from the outside, the will is subject to causality; looked at from the inside, or subjectively, it is free." What Planck concluded was that a powerful enough computer (actually, he used the word "observer"), endowed with a complete understanding of consciousness and *given all needed initial conditions at some instant*, could thereafter predict my exact behavior. However, reminiscent of the development of X, the computer cannot tell *me* what I *will* do. If the computer told me what I would have done had it *not* told me, I would surely do something different because of the computer's interaction with me: I am thus denied knowledge of my own future behavior. Planck was clearly ahead of his time: MacKay (1978) reached the same conclusions forty-one years later.

I have argued that I_0 represents the irreducible and multi-modal output (across the brain–consciousness interface) of my behavioral control system *sensors*. I could *not* have "done otherwise" than write this paper. *You* could not have done otherwise than read it. I do not impute to you any powers of "conscious control" of your behavior or thoughts any more than I impute such powers to myself. The statement that you or I "decided" or "behaved" or "thought" is a *manner of speaking* consistent with our subjective sense of free will. That statement does *not* imply "conscious control": the "you" and "I" must be interpreted in the public, conversational sense only. There is no reason not to continue that manner of speaking — it is an efficient means of communication — but I nevertheless maintain that unconscious processes in our brains determine what we think and feel and do, based solely on our heredity and environment.

References

Baars, B.J. (1997). *In the theater of consciousness*. New York: Oxford University Press.

Bunge, M. (1980). Emergence and the mind. In A.D. Smith, R. Llinas, and P.G. Kostyuk (Eds.), *Commentaries in the neurosciences* (pp. 633–641). New York: Pergamon Press. (Originally published in 1977)

Chalmers, D.J. (1996). *The conscious mind*. New York: Oxford University Press.

Cotterill, R. (1989). *No ghost in the machine*. London, England: Heinemann.

Crick, F. (1995). *The astonishing hypothesis*. New York: Simon and Schuster.

Daveney, T. (1964). Choosing. *Mind, 73*, 515–526.

Deikman, A.J. (1996). "I"=Awareness. *Journal of Consciousness Studies, 3*, 350–356.

Dennett, D.C. (1991). *Consciousness explained.* Boston: Little, Brown.
DiStefano, J.J., Stubberud, A.R., and Williams, I.J. (1990). *Feedback and control systems.* New York: McGraw–Hill.
Eccles, J.C. (1953). *The neurophysiological basis of mind.* Oxford, England: Clarendon Press.
Eccles, J.C., and Robinson, D.N. (1984). *The wonder of being human.* New York: Macmillan.
Elitzur, A.C. (1989). Consciousness and the incompleteness of the physical explanation of behavior. *The Journal of Mind and Behavior, 10,* 1–20.
Güzeldere, G. (1995). Consciousness: What it is, How to study it, What to learn from its history. *Journal of Consciousness Studies, 2,* 30–51.
James, W. (1950). *The principles of psychology* (Volume 1). New York: Dover Press. (Originally published in 1890)
James, W. (1984). *Psychology: Briefer course.* Cambridge, Massachusetts: Harvard University Press. (Originally published in 1892)
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8,* 529–566.
Libet, B., Gleason, C., Wright, E., and Pearl, D. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain, 106,* 623–642.
MacKay, D.M. (1978). What determines my choice? In P. Buser and A. Rougeul–Buser (Eds.), *Cerebral correlates of conscious experience* (pp. 335–345). New York: North–Holland.
Margenau, H. (1950). *The nature of physical reality.* New York: McGraw–Hill.
Marken, R.S. (1988). The nature of behavior: Control as fact and theory. *Behavioral Science, 33,* 196–206.
Natsoulas, T. (1996). The case for intrinsic theory: I. An introduction. *The Journal of Mind and Behavior, 17,* 267–285.
Penfield, W. (1975). *The mystery of the mind.* Princeton: Princeton University Press.
Perlis, D. (1997). Consciousness as self-function. *Journal of Consciousness Studies, 4,* 509–525.
Planck, M. (1937). *The universe in the light of modern physics.* London, England: George Allen and Unwin.
Ryle, G. (1984). *The concept of mind.* Chicago: University of Chicago Press. (Originally published in 1949)
Schrödinger, E. (1993). *What is life?* Cambridge, England: Cambridge University Press. (Originally published in 1944)
Searle, J. (1984). *Minds, brains and science.* Cambridge, Massachusetts: Harvard University Press.
Searle, J. (1994). *The rediscovery of the mind.* Cambridge, Massachusetts: The MIT Press.
Sommerhoff, G. (1996). Consciousness explained as an internal integrating system. *Journal of Consciousness Studies, 3,* 139–157.
Strawson, G. (1997). The self. *Journal of Consciousness Studies, 4,* 405–428.
Taylor, R. (1973). *Action and purpose.* Englewood Cliffs, New Jersey: The Humanities Press. (Originally published in 1966)
Velmans, M. (1996). An introduction to the science of consciousness. In M. Velmans (Ed.), *The science of consciousness* (pp. 1–22). London, England: Routledge.